

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 13:12:37 ; Search time 4574.73 seconds
(without alignments)
1190.031 Million cell updates/sec

Title: US-09-508-821b-2

Perfect score: 330
Sequence: 1 tccctcccccacccctccac.....ggccagggctactgcagcc 330

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 1472140 seqs, 824859755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_htgo_hum: *
31: em_htgo_inv: *
32: em_htgo_rnd: *
33: em_htg_hum: *
34: em_htg_inv: *
35: em_htg_rnd: *
36: em_htg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NC	Score	Query Match	Length	DB ID	Description
1	329	99.7	330	6 AX004677	AX004677 Sequence
2	320.6	97.2	542	11 G09710	G09710 human SRS C
3	317	96.1	5915	9 AB058723	AB058723 Homo sapi
4	317	96.1	6022	6 AX004680	AX004680 Sequence
5	317	96.1	160544	2 HS253P07	HS253P07
6	306	92.7	23843	9 HSA217191	HSA217191 Homo sapi
7	283.6	85.9	5395	6 AX024201	AX024201 Sequence
8	283.6	85.9	5667	9 HSA217190	HSA217190 Homo sapi
9	288.6	78.4	595	11 AF021115	AF021115 Homo sapi
10	170.2	51.6	328	6 AX004676	AX004676 Sequence
11	170.2	51.6	7222	10 X058A	D29801 Mouse mRNA
12	68.8	20.8	197232	2 AC073946	AC073946 Mus muscu
13	68.6	20.8	480	3 AF104706	AF104706 Plasmodu
14	67.8	20.5	1968	10 MMU06463	U46643 Mus musculu
15	67.2	20.4	5120	6 AX067148	AX067148 Sequence
16	67.2	20.4	5120	9 AF152102	AF152102 Homo sapi
17	67.2	20.3	461	3 AF104705	AF104705 Plasmodu
18	65.8	19.9	71803	2 AC010322	AC010322 Homo sapi
19	65.4	19.8	1547	10 MMU070649	U70649 Mus musculu
20	65.2	19.8	647	3 U91677	U91677 Plasmodiu
21	65.2	19.8	1512	10 MMU070655	U70655 Mus musculu
22	65.2	19.8	1512	10 MMU070656	U70656 Mus musculu
23	65.2	19.8	1546	10 MMU070645	U70645 Mus musculu
24	65.2	19.8	1549	10 MMU070657	U70657 Mus musculu
25	65.2	19.8	1586	10 MMU070653	U70653 Mus musculu
26	65.2	19.8	1598	10 MMU070654	U70654 Mus musculu
27	65.2	19.8	10266	22 E11536	E11536 DNA from Y
28	65.2	19.8	14625	10 MARSYLOC	X67204 Mus musculu
29	65	19.7	615	3 PFAMS42L	L19045 Plasmodiu
30	64.4	19.5	162869	9 AC090645	AC090645 Homo sapi
31	64.2	19.5	1611	10 MMU070651	U70651 Mus musculu
32	64.2	19.5	1612	10 MMU070652	U70652 Mus musculu
33	64	19.4	185306	2 AC013570	AC013570 Homo sapi
34	63.4	19.2	413	3 AF104711	AF104711 Plasmodu
35	63.4	19.2	471	3 AF104712	AF104712 Plasmodu
36	62.8	19.0	5719	10 MNCATSL	AF051726 Mus muscu
37	62.4	18.9	527	3 AF104713	AF104713 Plasmodu
38	62.4	18.9	8071	3 MSORTIRET	M93690 Anopheles g
39	62.2	18.8	460	3 AF329560	AF329560 Plasmodu
40	62.2	18.8	503	3 AF104707	AF104707 Plasmodu
41	61.8	18.7	1546	10 MMU070642	U70642 Mus musculu
42	61.6	18.7	1550	10 MMU070641	U70641 Mus musculu
43	61.6	18.7	1552	10 MMU070650	U70650 Mus musculu
44	61.6	18.7	1558	10 MMU070646	U70646 Mus musculu
45	61.6	18.7	2258	10 AF337043	AF337043 Mus muscu

ALIGNMENTS

RESULT 1
AX004677 330 bp DNA PAT 24-AUG-2000
LOCUS Sequence 2 from Patent WO9915639.
DEFINITION AX004677
ACCESSION AX004677.1 GI:9928113
VERSION
KEYWORDS
SOURCE
ORGANISM
unidentified.
unclassified.
REFERENCE 1 (bases 1 to 330)
AUTHORS Rouleau, G.A. and Joebey, R.
TITLE Polymorphic cag repeat-containing gene and uses thereof
JOURNAL Patent: WO 9915639-A 2 01-APR-1999;
ROUTEAU GUY A (CA); UNIV MCGILL (CA)
FEATURES
SOURCE Location/Qualifiers
1..330
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 72 a 128 c 83 g 46 t 1 others

Db	659	CGACTGCCAGGCCCATGACAGGCCGCTGAGATGCAGAGCTCCAGAGCTGCCCCCGGGGCAGC	718
QY	144	gggtccagagaattcttcactgcctaccagtcggcgccgcctcagctatgacacagcagcagc	203
Db	719	GGGTCCAGAAATCTTCATGCTACCAAGTCGGCGCGCTCAGCTATGACACAGCAGCAGCAGC	778
QY	204	agcagcagcagcagcagcagcagcagcccttcagcgcgcgcgcacatgccccaggaaa	263
Db	779	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCTTCAGAGCGGCGCACCATGCGCCAGGAA	838
QY	264	cctcctaatatcaaaacctcgcacagttatcagctacgcgcagcagcagcagcagcact	323
Db	839	CCCTCCATTACCAAAACCTCCCAAGTATCAGACACTACGGGAGCGAAGCCAGGGCTACT	898
QY	324	gccagcc	330
Db	899	GCCAGCC	905

RESULT	8	
LOCUS	HSA271790	5667 bp mRNA
DEFINITION	Homo sapiens mRNA for retinoid-acid induced protein 1 (RAIL gene).	
ACCESSION	AJ271790	
VERSION	AJ271790.1	GI:12053792
KEYWORDS	RAIL gene; retinoid-acid induced protein 1.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	Seranski, P., Hoff, C., Radelof, U., Hennig, S., Reinhardt, R., Schwartz, C.E., Heiss, N.S. and Poustka, A.	
TITLE	RAIL is a novel polyglutamine encoding gene that is deleted in Smith-Magenis syndrome patients	
JOURNAL	Gene 270 (1-2), 69-76 (2001)	
MEDLINE	21297181	
REFERENCE	2 (bases 1 to 5667)	
AUTHORS	Seranski, P.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-JAN-2000) Seranski P., Molecular Genome Analysis, Deutsches Krebsforschungszentrum Heidelberg, Im Neuenheimer Feld 280, 69120 Heidelberg, GERMANY	

QY	264	ccctccattaccacaaacctcgcccaagtatcagcactacggcagcaaggccagggtact	323
Db	839	CCCTCCATTACCAAAACCTCGCCAGTAGTCAGCAGCTACGGGACAGGCGAGGCGCTACT	898
QY	324	gccagcgc 330	
Db	899	GCCAGCC 905	
RESULT	9		
LOCUS	AF021115/c		
DEFINITION	Homo sapiens trinucleotide repeat ctg-23, sequence tagged site.	STS	03-APR-1998
ACCESSION	AF021115		
VERSION	AF021115.1	GI:3015598	
KEYWORDS	STS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 595)		
AUTHORS	Philibert,R.A., Horelli-Kuitunen,N., Robb,A.S., Lee,Y.H., Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B., Palotie,A. and Ginns,E.I.		
TITLE	The characterization and sequence analysis of thirty CTG-repeat containing genomic cosmid clones		
JOURNAL	Eur. J. Hum. Genet. 6 (1), 89-94 (1998)		
MEDLINE	98454394		
REFERENCE	2 (bases 1 to 595)		

[illegible][illegible]

RESULT	12			
AC073946/C				
LOCUS	AC073946	19732 bp	DNA	HTG
				16-JAN-2001

DEFINITION	Mus musculus chromosome 10 clone RP23-218e4, WORKING DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION	AC073946
VERSION	GI:12232501
KEYWORDS	HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
AUTHORS	Fu, Y., Wang, Q., Pan, H. and Roe, B.A.
TITLE	1 (bases 1 to 197233)
JOURNAL	Mus musculus Chromosome 10 BAC Clone rp23-218e4
REFERENCE	2 (bases 1 to 197232)
AUTHORS	Fu, Y., Wang, Q., Pan, H. and Roe, B.A.
TITLE	Direct Submission
JOURNAL	Submitted (07-JUL-2000) Department of Chemistry and Biochemistry, The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT	On Jan 16, 2001 this sequence version replaced gi:11641343.

OK 73013, USA
On Jan 16, 2001, this sequence replaced gi:11641343.
COMMENT

COMMENT: On Jan 18, 2001 this sequence accession record was generated.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

*	2831:	contig of 2831 bp in length
*	2832	2931: gap of unknown length
*	2932	15879: contig of 12948 bp in length
*	15880	15979: gap of unknown length
*	15980	33201: contig of 17222 bp in length
*	33202	33301: gap of unknown length
*	33302	51575: contig of 18274 bp in length
*	51576	51675: gap of unknown length
*	51676	81126: contig of 29451 bp in length
*	81127	81226: gap of unknown length
*	81227	113588: contig of 32362 bp in length
*	113589	113688: gap of unknown length
*	113689	197233: contig of 83544 bp in length.

FEATURES	source	Location/Qualifiers
		1. 197232
		/db_name="Mus musculus"
		/db_xref="taxon:10090"
		/chromosome="10"
		/clone="RP23-218e4"
BASE COUNT	49202 a 48272 c 48817 g 50340 t	
ORIGIN		601 others

Query Match 20.8%; Score 68.8; DB 2; Length 197232;
Best Local Similarity 56.7%; Pred. No. 0.00016;
Matches 127; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

92 cagcccatgacaggccgctgactgccagctccagctggccccggggcagcgggtccag 151

D6 81609 CAGACCCAGCACCAAGCACCAGCAGCAGCACCACCAACAGAGCACCAC

A: 150 rats at least were used in each experiment.

91490

[illegible]

Qy 212 cagcagcagcagcagcagcccttcagagccgaccatgccagggaacctccat 2/1

DB- 81489 CACCAGCACCAGCTGCAGCACCACCGAGTGCAGCACCACCGAGCACCACCAC

Qy 272 taccaaacctgcgaagtatcagcactacgggagcaaggcca 315

Dd 81429 CAGCACCAGACCCACCAGCACCAGCACCCAGCACCAGTACCA 81386

RESULT 13

100

Search completed; December 4, 2001, 16:09:27
Job time: 10610 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 14:15:07 ; Search time 278.68 Seconds
(without alignments)
1015.205 Million cell updates/sec

Title: US-09-508-821b-2

Perfect score: 330
Sequence: 1 tccctcccccctcctccac.....ggccagggctacgcagcc 330

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.1101.*
1: /SIDS2/gcgcdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgcdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgcdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgcdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgcdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgcdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgcdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgcdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgcdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgcdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgcdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgcdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgcdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgcdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgcdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgcdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgcdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgcdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgcdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgcdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgcdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgcdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	317	96.1	6022	20	AAAX23983 Human hGTL DNA. H
2	283.6	85.9	5395	21	AAZ88975 Human HSGT1 CDNA.
3	166.8	50.5	250	21	AAAS7095 Human colon cancer
4	112	33.9	1004	20	AAAX23982 Human GCT10D4 DNA
5	67.2	20.4	5120	22	AAAC84677 DNA sequence of hu
6	65.2	19.8	10266	17	AAAT33007 Mouse SVT-related
7	61.6	19.8	14704	23	AAQ20685 PMS 741 insert con
8	61.6	18.7	7680	22	AAI35276 Human polyomuclei
9	60.6	18.4	1159	21	AAAS9240 An Ecoli fragment
10	59.4	18.0	234	16	AAQ84832 Spinocebellar at
11	59.4	18.0	397	20	AAAX89891 Spinocebellar at

12	59.4	18.0	477	21	AAZ4307 Human SCAT genomic
13	59	17.9	1037	21	AAAS9242 Exons E, C and A o
14	59	17.9	1472	21	AAAS9241 Exons D, C, B and
15	58.8	17.8	203	19	AAAV0271 Glutamine rich reg
16	58.8	17.8	203	19	AAAV1226 SCA2 gene CAG repe
17	58.2	17.6	3254	19	AAV32402 Drosophila melanoa
18	56.6	17.2	154	16	AAQ84835 Spinocebellar at
19	56.4	17.1	171	16	AAQ84834 Spinocebellar at
20	56.4	17.1	195	16	AAQ84831 Spinocebellar at
21	56.2	17.0	168	16	AAQ84833 Spinocebellar at
22	55.6	16.8	401	22	AAI14063 Probe #396 for ge
23	55.6	16.8	401	22	AAI35443 Probe #4129 used t
24	55.6	16.8	401	22	AAI33276 Probe #17283 used t
25	55.6	16.8	446	22	AAI48556 Probe #8912 used t
26	55.6	16.8	446	22	AAI48556 Probe #8912 used t
27	55.6	16.8	446	22	AAI48556 Probe #8912 used t
28	55.6	16.8	1635	22	AAI22780 Probe #16768 used
29	55.6	16.8	1635	22	AAI22780 Probe #16768 used
30	55.6	16.8	1635	22	AAI22780 Probe #16768 used
31	55.6	16.8	1635	22	AAI22780 Probe #16768 used
32	55.6	16.8	1635	22	AAI22780 Probe #16768 used
33	55.6	16.8	1635	22	AAI22780 Probe #16768 used
34	55.4	16.8	476	22	AAI10073 Probe #7 used to m
35	55.4	16.8	476	22	AAI10073 Probe #7 used to m
36	55.4	16.8	476	22	AAI10073 Probe #7 used to m
37	55.2	16.7	1075	22	AAI21781 Probe #15750 used
38	55.2	16.7	1075	22	AAI47064 Probe #7457 used t
39	55.2	16.7	1075	22	AAI07466 Probe #2526 for ge
40	55.2	16.7	1403	22	AAI12593 Probe #2627 used t
41	55.2	16.7	1403	22	AAI33941 Probe #2489 used t
42	55.2	16.7	1403	22	AAI02498 Probe #2489 used t
43	55	16.7	530	21	AAQ55931 Eucalyptus grandis
44	54.8	16.6	486	22	AAQ75507 Polyglutamine trac
45	54.8	16.6	1848	22	AAQ50286 DNA sequence of an

ALIGNMENTS

RESULT 1

AAAX23983
ID AAAX23983 standard; DNA; 6022 BP.

XX AAAX23983;
XX 25-JUN-1999 (first entry)

XX Human hGTL DNA.

XX Polymorphic CAG repeat; hGTL; diagnosis; prognosis; schizophrenia;
XX human; transcription factor; neuroleptic activity; affective disorder;
XX manic depression; neurodevelopmental brain disease; detection;
XX phenotypic variability; ss.

XX Homo sapiens.

XX W0915639-A1.

XX J1-APR-1999.

XX I3-SEP-1998; 98WC-CA00884.

XX I3-SEP-1997; 97CA-2216057.

XX (UWMC-) UNIV MCGILL.

XX Benkelfat C, Joobier R, Rouleau GA;

XX WPI: 1999-254703/21.

XX A human Gtl gene containing a transcribed polymorphic CAG repeat,

XX useful in the diagnosis and treatment of schizophrenia

PR 20-NOV-1998; 98US-0196716.

XX (LUDM-) LUDMIG INST' CANCER RES.

XX Simpson AJG, Dias Neto E, Brentani RB;

XX WPI, 2000-400100/34.

XX Method for determining open reading frames of the genome of an organism
PT using primers at low stringency conditions, useful in the construction
PT of contigs or constructs of sequenced nucleic acid molecules -

XX Example 6; Page 70; 113pp; English.

CC The present sequence is a cDNA sequence obtained using a method for
CC determining open reading frames (ORFs) of the genome of an
CC organism. An aliquot of mRNA from human colon cancer cells was mixed
CC with a single, arbitrary primer, Moloney murine leukemia virus reverse
CC transcriptase, reverse transcriptase buffer and dNTPs. The mixture was
CC incubated under low stringency conditions to yield single stranded
CC cDNA. The same primer was then used to amplify the cDNA by PCR. Rather
CC than providing nucleotide sequence information from the non-coding
CC termini of nucleic acid molecules, the method provides information on
CC the more interesting and relevant internal portions, such as ORFs. The
CC method also permits the construction of contigs of sequenced nucleic
CC acid molecules.

XX Sequence 250 BP; 66 A; 99 C; 56 G; 29 T; 0 other;

Query Match 50.5%; Score 166.8; DB 21; Length 250;

Best Local Similarity 97.3%; Pred. No. 6e-27;

Matches 181; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

QY 145 gttccagaattcttcgtccaccagtcggtcgccgctcagctatgacccagcagcagca 204
DB 1 gttccagaattcttcgtccaccagtcggtcgccgctcagctatgacccagcagcagca 57
QY 205 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 264
DB 58 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 117
QY 265 gctccattccaaacccctcccaagratagactagggcagcagcagcagcagcagc 324
DB 118 cctccattccaaacccctcccaagratagactagggcagcagcagcagcagcagcagc 177
QY 325 ccaagcc 330
DB 178 ccaagcc 183

RESULT 4

AAK23982 AAK23982 standard; DNA; 1004 BP.

AC AAK23982;

DT 25-JUN-1999 (first entry)

DE Human GCT10D04 DNA.

XX Polymorphic CAG repeat; hGT1; diagnosis; prognosis; schizophrenia;
KW human; transcription factor; neuroleptic activity; affective disorder;
KM manic depression; neurodevelopmental brain disease; detection;
KM phenotypic variability; ss.

OS Homo sapiens.

PN WO915639-A1.

PD 01-APR-1999.

PF 18-SEP-1998; 98MO-CA00884.

XX 19-SEP-1997; 97CA-2216057.

XX (UJMC-) UNIV MCGILL.

XX Benkefat C, Joober R, Rouleau GA;

XX WPI, 1999-254703/21.

XX A human GT1 gene containing a transcribed polymorphic CAG repeat,
PT useful in the diagnosis and treatment of schizophrenia

XX Claim 1; Fig 3A-B; 41pp; English.

CC This invention describes novel human GT1 (hGT1) transcription factor
CC gene with neuroleptic activity containing a transcribed polymorphic
CC CAG repeat. Allelic variants of the hGT1 gene CAG repeat are associated
CC with schizophrenia, affective disorders (especially manic depression),
CC neurodevelopmental brain diseases or with phenotypic variability with
CC respect to long term response to neuroleptic medication. Short
CC (171-177 bp) allelic variants of CAG repeats in the hGT1 gene, are
CC indicative of non-severe schizophrenia and neuroleptic response in
CC patients. Probes and/or primers designed using the hGT1 gene can be
CC used to identify genes interacting with a biochemical pathway affected
CC by the hGT1 gene. The identified gene role can then be evaluated in
CC psychiatric patients. Therapeutic agents can be identified by
CC administering the agent to a transgenic mammal (or schizophrenic
CC patients) and evaluating the prevention and/or treatment of development
CC of schizophrenia.

XX Sequence 1004 BP; 220 A; 366 C; 243 G; 136 T; 39 other;

Query Match 33.3%; Score 112; DB 20; Length 1004;

Best Local Similarity 72.3%; Pred. No. 2.8e-15;

Matches 133; Conservative 8; Mismatches 43; Indels 0; Gaps 0;

QY 147 tccagaattcttcgtccaccagtcggtcgccgctcagctatgacccagcagcagcagc 206
DB 739 tccagaattcttcgtccaccagtcggtcgccgctcagctatgacccagcagcagcagc 798
QY 207 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 266
DB 799 tccagaattcttcgtccaccagtcggtcgccgctcagctatgacccagcagcagcagc 858
QY 267 tccattccaaacccctcccaagratagactagggcagcagcagcagcagcagcagc 326
DB 859 tccattccaaacccctcccaagratagactagggcagcagcagcagcagcagcagcagc 918
QY 327 agcc 330
DB 919 agcc 922

RESULT 5

AAC84677 AAC84677 standard; DNA; 5120 BP.

AC AAC84677;

DT 20-APR-2001 (first entry)

DE DNA sequence of human PPP2R2B.

XX Microsatellite marker; human; chromosome 5q31-33; D5S402; WI-6763;
KW CAG trinucleotide repeat; spinocerebellar ataxia 12; SCA12; PPP2R2Bbeta;
KW brain; protein phosphatase PP2A; ds.

OS Homo sapiens.

PN

PF

key Location/Qualifiers
repeat_region 2088..2366
/*tag= a

/rpt_type= "CAG"
2473..3093
/*tag= b
5'UTR 2473..3023
/*tag= c
CDS 3024..3093
/*tag= d
/product= "PP2R2B"
/note= "protein_id= AAF74024.1"

WO200078943-A2.
28-DEC-2000.
16-JUN-2000; 2000WO-US40213.
18-JUN-1999; 99US-0140176,
(UJVO) UNIV JOHNS HOPKINS.
Margolis RL, Ross CA, Holmes SE;
WPI; 2001-102718/11.
P-PSEDE; AAB48340.

New polynucleotide for use in diagnosing spinocerebellar ataxia 12,
comprises a microsatellite marker having a variable number of CAG
trinucleotide repeats located on human chromosome 5q31-33 -

Disclosure; Page 21-23; 23pp; English.

The invention relates to a microsatellite marker located on human
chromosome 5q31-33 between markers D5S402 and WI-6763, comprising a
variable number of CAG trinucleotide repeats. The marker is useful for
diagnosing spinocerebellar ataxia 12 (SCA12) by determining the number of
CAG trinucleotide repeats in a 5'-untranslated region of an allele of
PPP2R2Bbeta, a brain specific regulatory subunit of protein phosphatase
PP2A, where if a number greater than 40 is determined, an expansion which
is associated with SCA12 is indicated. The presence of expanded
trinucleotide repeats on chromosome 5q31-33 is useful for predicting or
diagnosing SCA12. Also provided are primers useful for amplifying the
microsatellite marker associated with SCA12 and for diagnosing SCA12 and
a probe useful for detecting expansions in a microsatellite marker
located on human chromosome 5q31-33. The present sequence represents
a PPP2R2B DNA sequence.

Sequence 5120 BP; 1336 A; 1267 G; 1268 G; 1249 T; 0 other;

Query Match	20.4%;	Score 67.2;	DB 22;	Length 5120;
Best Local Similarity	56.9%;	Prod. NO. 1e-05;		
Matches 123;	Conservative	0;	Mismatches 93;	Indels 0; Gaps 0;
114	ctgcagctccagcctgccccgggcagcgggtccagaattctcctaccagtcgg	173		
2068	cagcgccctccagctctctgcagcagcagcagcagcagcagcagcagcagc	2127		
174	gcgcctcagctatgacgcagcagcagcagcagcagcagcagcagcagcagc	233		
2128	agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2187		
234	cccttcagcgcggccacatgcccagaaacctcattaccaaaacctcgccaagt	293		
2188	agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2247		
294	agcactcgggcagcagcagcagcagcagcagcagcagcagcagcagcagc	329		
2248	agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2283		

RESULT 6
AAT33007
ID AAT330

[illegible]

Query_Match	19.8%	Score	65.2	DB	17	Length	10266	
Best_Local_Similarity	54.6%	Pred. No.	3e-05					
Matches	130	Conservative	0	Mismatches	108	Indels	0	Gaps
Qy	92	caacccatgacagccgcgtgactgcagcttcagccctggccggggcagcagcggtccag	151					
Db	7838	cagcagcagcagcagcagcagcagcagtcctcattgacacccaccagaaagcagcaggtccat	7897					
Qy	152	aattcttcattgcctaccagctgggcgcgcctcagctatgaccagcagcagcagcagcagc	211					
Db	7898	gaccacaccaccacacacacagcagcagcagtcctcattgaccaccagcagcagcagcagc	7957					
Qy	212	cagcagcagcagcagcagcagcagccttcagcgcgcgcaccatgccagaaaccctccat	271					
Db	7958	ttccattgaccacccgcgcgcagcagcattcctatgaccaccccccagcagcagcagc	8017					
Qy	272	taccacaaacctgcacagctatcagcactacgggcagcagcagggcctactgcgcgc	329					
Db	8018	tctcattaccacccacacacacagcagtccttcattgaccaccaccacacagcagcagc	8075					

RESULT 7
AAQ20685
ID AAQ20685 standard; DNA; 14704 BP.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: December 4, 2001, 14:13:57 ; Search time 102.2 seconds
(without alignments)
731.289 Million cell updates/sec

Title:

US-09-508-821B-2

Perfect score:

330
1 tcttcccccacacccctccacac.....ggccagggcctactccagcc 330

Sequence:

IDENTITY_NUC

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched:

351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters:

702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database:

Issued_Patents_NA.*
1: /cgn2.6/prodata/2/ina/5A.COMB.seq.*
2: /cgn2.6/prodata/2/ina/5B.COMB.seq.*
3: /cgn2.6/prodata/2/ina/5A.COMB.seq.*
4: /cgn2.6/prodata/2/ina/5B.COMB.seq.*
5: /cgn2.6/prodata/2/ina/PCRUS.COMB.seq.*
6: /cgn2.6/prodata/2/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	59.4	18.0	234	1	US-08-469-802B-3	Sequence 3, Appl
2	59.4	18.0	234	2	US-08-267-803B-3	Sequence 3, Appl
3	59.4	18.0	397	3	US-09-253-691-3	Sequence 3, Appl
4	59.4	18.0	477	4	US-09-135-994-1	Sequence 1, Appl
5	58.8	17.8	203	4	US-09-043-303-7	Sequence 7, Appl
6	58.2	17.6	3254	3	US-08-965-903B-1	Sequence 1, Appl
7	56.6	17.2	154	1	US-08-469-802B-6	Sequence 6, Appl
8	56.6	17.2	154	2	US-08-267-803B-6	Sequence 6, Appl
9	56.4	17.1	171	1	US-08-469-802B-5	Sequence 5, Appl
10	56.4	17.1	171	2	US-08-267-803B-5	Sequence 5, Appl
11	56.4	17.1	195	2	US-08-469-802B-2	Sequence 2, Appl
12	56.4	17.1	195	2	US-08-267-803B-2	Sequence 2, Appl
13	56.2	17.0	168	1	US-08-469-802B-4	Sequence 4, Appl
14	56.2	17.0	168	2	US-08-267-803B-4	Sequence 4, Appl
15	54.6	16.5	165	4	US-09-043-303-17	Sequence 17, Appl
16	54.4	16.5	2793	1	US-08-458-298-1	Sequence 1, Appl
17	54.4	16.5	2793	1	US-08-209-747-1	Sequence 1, Appl
18	53.4	16.2	3366	2	US-08-469-802B-1	Sequence 1, Appl
19	53.4	16.2	3366	2	US-08-267-803B-1	Sequence 1, Appl
20	53.4	16.2	10660	4	US-08-267-803B-8	Sequence 8, Appl
21	53.4	16.2	10660	4	US-09-041-886-16	Sequence 16, Appl
22	53.2	16.1	506	1	US-08-469-802B-7	Sequence 7, Appl
23	53.2	16.1	506	2	US-08-267-803B-7	Sequence 7, Appl
24	53	16.1	3376	3	US-08-320-559-29	Sequence 29, Appl
25	53	16.1	3376	3	US-08-545-860D-29	Sequence 29, Appl
26	53	16.1	3376	5	PCR-US94-04496-29	Sequence 29, Appl
27	51.8	15.7	10348	2	US-08-457-273B-41	Sequence 41, Appl

28	51.8	15.7	10348	3	US-08-556-419-13	Sequence 13, Appl
29	51.8	15.7	10348	4	US-08-041-886-14	Sequence 14, Appl
30	51.8	15.7	10366	1	US-08-246-982B-5	Sequence 5, Appl
31	51.8	15.7	10366	1	US-08-453-265-5	Sequence 5, Appl
32	49.6	15.0	543	6	5273901-6	Patent No. 5273901
33	49.6	15.0	3489	2	US-08-728-323A-1	Sequence 1, Appl
34	49.6	15.0	32207	2	US-08-770-379-20	Sequence 20, Appl
35	49.6	15.0	32207	2	US-08-757-669A-20	Sequence 20, Appl
36	49.4	15.0	32207	3	US-08-330-970-2	Sequence 2, Appl
37	49.4	15.0	3336	3	US-09-330-970-4	Sequence 4, Appl
38	49.2	14.9	3715	4	US-09-041-886-10	Sequence 10, Appl
39	49	14.8	4835	1	US-08-386-495-9	Sequence 9, Appl
40	49	14.8	4835	5	PCR-US96-02331-9	Sequence 9, Appl
41	48.2	14.6	1776	2	US-08-531-927B-1	Sequence 1, Appl
42	48.2	14.6	1776	4	US-09-041-886-12	Sequence 12, Appl
43	48.2	14.6	2950	5	PCR-US93-08386-7	Sequence 7, Appl
44	48.2	14.6	2968	5	PCR-US93-08386-1	Sequence 1, Appl
45	48.2	14.6	3771	1	US-08-185-432-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-469-802B-3
Sequence 3, Application US/08469802B
Patent No. 5741645
GENERAL INFORMATION:
APPLICANT: Orr, Harry T.
APPLICANT: Ratum, Laura P.W.
APPLICANT: Chung, Ming-Yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
Patent No. 5741645
TITLE OF INVENTION: Type 1 and Method for Diagnosis
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESS: Mueller, Raasch, Gebhardt & Schnappach, P.A.
STREET: 119 No. 5741645th Fourth Street, Suite 203
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/469,802B
APPLICATION NUMBER: US/08/469,802B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mueller, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00030101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1225
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-469-802B-3
Query Match 18.0%; Score 59.4; DB 1; Length 234;
Best Local Similarity 56.3%; Pred. No. 1.7e-05;
Matches 11; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

	Query Match	18.0%	Score 59.4;	DB 2;	Length 234;
	Best Local Similarity	56.3%;	Pred. No. 1.7e-05;		
	Matches 111;	Conservative 0;	Mismatches 86;	Indels 0;	Gaps 0;
QY	134	ccggggcagcgggttcagataattcattgcataccagtcggcgccctcagctatgaccag	193		
Db	25	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	84		
QY	194	cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagccat	253		

9-135
293

Query Match	18.0%;	Score 59.4;	DB 4;	Length 477;
Best Local Similarity	55.6%;	Pred. No. 2e-05;		
Matches 114;	Conservative	0;	Mismatches 91;	Indels 0;
				Gaps 0;

[illegible]

RESULT 5
US-09-043-303-7
; Sequence 7, Application US/09043303

```

? APPLICANT: TSUI, Shoji
? APPLICANT: SAMPEL, Kazuhiro
? TITLE OF INVENTION: Method for diagnosing Spinoecerebellar Ataxia Type 2 and
? TITLE OF INVENTION: Primes Theracor
? FILE REFERENCE: 0760-0241P
? CURRENT APPLICATION NUMBER: US/09/043,303
? CURRENT FILING DATE: 1998-05-18
? EARLIER APPLICATION NUMBER: PCT/JP96/01999
? EARLIER FILING DATE: 1996-07-18
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO: 7
? LENGTH: 203
? TYPE: DNA
? ORGANISM: p-2093 plasmid
? OS-09-043-303-7

```

Query Match	17.8;	Score 58.8;	DB 4;	Length 203;
Best Local Similarity	56.1;	Pred. No. 2.3e-05;		
Matches 11;	Conservative	0;	Mismatches 87;	Indels 0;
				Gaps 0;

QY	133	ccacgggagacggggtctcaaaatacttcaatgcctacacagctacagctggcgcgcctcctcagctctatgacaa	132
Db	6	ccaaaggaaataagaaatacaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca	65
QY	193	gcaagcagaagcagaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca	252
Db	66	gcaagcagaagcagaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca	122
QY	253	tgcaccagggaaaaccctctcaattacaaataaccctcgcgaagatacgaactaagggcagcagcaag	312
Db	126	gcaagcagaagcagaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagcaagca	185
QY	313	ccaaagggtctactggccacgc	330
Db	186	tcaacggaaactctgggcgc	203

RESULT 6
US-08-965-903B-1
; Sequence 1, Application US/08965903B
; Patent No. 6060275
; GENERAL INFORMATION:
; APPLICANT: Haochen, Nlr

APPLICANT: Krasnow, Mark A.
TITLE OF INVENTION: SPROUTY PROTEIN AND CODING SEQUENCE
TITLE OF INVENTION: SPROUTY PROTEIN AND CODING SEQUENCE
TITLE OF INVENTION: SPROUTY PROTEIN AND CODING SEQUENCE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Query Match	17.6%;	Score 58.2;	DB 3;	Length 3254;
Best Local Similarity	59.3%;	Pred. No. 5.4e-05;		
Matches 99;	Conservative 0;	Mismatches 68;	Indels 0;	Gaps 0;

[illegible]

US-08-469-802B-6
Sequence 6, Application US/08469802B
Patent No. 5711645
GENERAL INFORMATION:
APPLICANT: Orr, Harry T.
APPLICANT: Rannum, Laura P.W.
APPLICANT: Chung, Ming-yi
APPLICANT: Zoghbi, Huda Y.

```

1 ZIP: 55458-1415
2 COMPUTER READABLE FORM:
3 * MEDIUM TYPE: Floppy disk
4 * COMPUTER: IBM PC Compatible
5 * OPERATING SYSTEM: PC-DOS/MS-DOS
6 SOFTWARE: PatentIn Release #1.0, Version #1.25
7 CURRENT APPLICATION DATA:
8 * APPLICATION NUMBER: US/08/267,803B
9 * FILING DATE: 28-JUN-1994
10 CLASSIFICATION: 435
11 ATTORNEY/AGENT INFORMATION:
12 * NAME: McCormack, Myra. H.
13 REGISTRATION NUMBER: 36,502
14 REFERENCE/DOCKET NUMBER: 110.00030120
15 TELECOMMUNICATION INFORMATION:
16 * TELEPHONE: 612-305-1217
17 * TELEFAX: 612-305-1228
18 INFORMATION FOR SEQ ID NO: 6:
19 SEQUENCE CHARACTERISTICS:
20     .LENGTH: 154 base pairs
21     .TYPE: nucleic acid
22     .STRANDEDNESS: single
23     .TOPOLOGY: linear
24     .MOLECULE TYPE: DNA
25 US-08-267-803B-5

```

NAME: McCormack, Myra H.
REGISTRATION NUMBER: 36.602
REFERENCE/DOCKET NUMBER: 110.00030120
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 154 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-267-803B-6

D6	61	GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	12
QY	307	gcaagggccagggtactatgcacg	329
D6	121	GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	143

RESULT⁹

US-08-469-802B-5
; Sequence 5, Application US/08469802B
Patent No. 5741645.
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi

Patent NO. 5741645
 TITLE OF INVENTION: Type 1 and Method for Diagnosis
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Muehling, Raasch, Gebhardt & Schwappach, P.A.
 STREET: 119 No. 5741645th Fourth Street, Suite 203
 CITY: Minneapolis
 STATE: MN
 COUNTRY: USA
 ZIP: 55401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/469,802B
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Muehling, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00030101
TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217
TELEFAX: 612-305-1225

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 171 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-469-802B-5

Query Match

17.1%; Score 56.4; DB 1; Length 171;

Best Local Similarity 59.3%; Pred. No. 7.7e-05;

Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Db

134 ccggggcagcggggtccatcttcacgagcggcgcgcctcagctatgacag 193

Db

1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

QY

194 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 253

Db

61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

QY

254 gccacggaacccctccattccaaaacctgcgaagatcag 295

Db

121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 162

RESULT 10

US-08-267-803B-5

Sequence 5, Application US/08267803B

Patent No. 5834183

GENERAL INFORMATION:

APPLICANT: Ott, Harry T.

APPLICANT: Rannum, Laura P.W.

APPLICANT: Chung, Ming-Yi

APPLICANT: Zoghbi, Huda Y.

TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

Patent No. 5834183

TITLE OF INVENTION: Type 1 and Method for Diagnosis

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Muehling, Raasch, Gebhardt & Schwappach, P.A.

STREET: P.O. Box 581415

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55458-1415

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/267,803B

FILING DATE: 28-JUN-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: McCormack, Myra H.

REGISTRATION NUMBER: 36,602

REFERENCE/DOCKET NUMBER: 110,00030120

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 171 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-267-803B-5

Query Match

17.1%; Score 56.4; DB 2; Length 171;

Best Local Similarity 59.3%; Pred. No. 7.7e-05;

Matches 96; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Db

134 ccggggcagcggggtccatcttcacgagcggcgcgcctcagctatgacag 193

Db

1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

QY

194 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 253

Db

61 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120

QY

254 gccacggaacccctccattccaaaacctgcgaagatcag 295

Db

121 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 162

RESULT 11

US-08-469-802B-2

Sequence 2, Application US/08469802B

Patent No. 5741645

GENERAL INFORMATION:

APPLICANT: Ott, Harry T.

APPLICANT: Rannum, Laura P.W.

APPLICANT: Chung, Ming-Yi

APPLICANT: Zoghbi, Huda Y.

TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

Patent No. 5741645

TITLE OF INVENTION: Type 1 and Method for Diagnosis

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Muehling, Raasch, Gebhardt & Schwappach, P.A.

STREET: 119 No. 5741645th Fourth Street, Suite 203

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,802B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Muehling, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110,00030101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1225

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 195 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-469-802B-2

Query Match

17.1%; Score 56.4; DB 1; Length 195;

Best Local Similarity 58.2%; Pred. No. 7.9e-05;

Matches 99; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 4, 2001, 12:59:46 ; Search time 2862.31 seconds

(without alignments)
1238.897 Million cell updates/sec

Title: US-09-508-821B-2

Perfect score: 330
Sequence: 1 tctctccacacactctctccac.....ggccagggcactgacagcc 330

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estipl:*
6: em_estiba:*
7: em_estrov:*
8: em_estrov:*
9: em_hic:*
10: gp_est1:*
11: gp_est2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	133.4	40.4	963	10	AI893920 m85a07.y
2	128.2	38.8	214	11	BG990415 MR2-RT116
3	86.4	26.2	224	10	AM492162 UT-M-BH3-
4	86.4	26.2	452	10	BE651204 UT-M-BH3-
5	69.4	21.0	226	13	AZ837781 ZM013011
6	68.8	20.8	805	13	CNS04RM2
7	65.4	19.8	887	12	AK013560 Mus muscu
8	64	19.4	984	13	CNS02NM2
9	63.4	19.2	831	10	BE705571
10	63.2	19.0	605	13	AZ640388
11	62.8	19.0	246	13	AZ593051
12	62.4	18.9	1021	13	CNS01G7A

c 13	61.6	18.7	1019	13	CNS01OY	AL154154 Anopheles
c 14	61	18.5	452	10	AJ283665	AA3B-AAE-
c 15	61	18.5	526	10	AI257238	AI257238 LP0546-E
c 16	60.4	18.3	526	13	AZ312601	AZ312601 LM0028104
c 17	60.2	18.2	572	13	CNS043SN	AL273200 Tetradon
c 18	60	18.2	628	11	BE630444	BE630444 602766990
c 19	60	18.2	2199	10	BE741896	BE741896 601595356
c 20	59.4	18.0	356	10	AV627762	AV627762 AV627762
c 21	59.4	18.0	492	10	AM465310	AM465310 BP230018B
c 22	59	17.9	1003	13	CNS04OZ	AL303092 Tetradon
c 23	58.8	17.8	555	13	AZ281461	AZ281461 RPT-23-1
c 24	58.8	17.8	562	13	AZ251135	AZ251135 RPT-23-5
c 25	58.8	17.8	806	13	CNS01JUL	AL109419 Drosophila
c 26	58.8	17.8	986	13	CNS021IL	AL199074 Tetradon
c 27	58.6	17.8	495	13	AO189056	AO189056 HS-195-B
c 28	58.6	17.8	1007	13	CNS01M2O	AL151699 Anopheles
c 29	57.8	17.5	947	13	CNS01137	AL146136 Anopheles
c 30	57.6	17.5	594	13	AZ870327	AZ870327 ZM0182613
c 31	57.6	17.5	980	13	CNS01BHD	AL154817 Anopheles
c 32	57.6	17.5	982	13	CNS06R3C	AL411922 T3 end of
c 33	57.6	17.5	1096	11	BG173272	BG173272 602336762
c 34	57.6	17.5	1101	13	CNS016KS	AL106870 Drosophila
c 35	57.4	17.4	264	11	BE798211	BE798211 601585978
c 36	57.4	17.4	590	13	CNS03M2Z	AL250244 Tetradon
c 37	57.4	17.4	932	13	CNS01B9Z	AL141082 Anopheles
c 38	57.2	17.3	723	13	AZ627893	AZ627893 IM0463118
c 39	57	17.3	680	10	AI388163	AI388163 GH18971.5
c 40	56.8	17.2	351	13	FR0004775	Z88558 F.rubripes
c 41	56.8	17.2	526	13	CNS03CMO	AL237969 Tetradon
c 42	56.8	17.2	540	10	AM464269	AM464269 BP230015A
c 43	56.6	17.2	752	10	AA940845	AA940845 LD23113.5
c 44	56.6	17.2	927	10	CNS01JAS	AL146589 Anopheles
c 45	56.6	17.2	1002	13	CNS01N61	AL151835 Anopheles

ALIGNMENTS

RESULT 1
AI893920 963 bp mRNA EST 15-MAR-2000
m85a07.y1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA
Clone IMAGE:355284 5' similar to gb:D9801 Mouse mRNA for unknown
product, complete cds (MUSE);, mRNA sequence.

ACCESSION AI893920
VERSION
KEYWORDS
SOURCE

ORGANISM
house mouse.
Mus musculus

REFERENCE
AUTHORS
Marras M., Hillier L., Kucaba T., Martin J., Beck C., Wylie T.,
Underwood K., Steptoe M., Theising B., Allen M., Bowers Y., Person
B., Swaller T., Gibbons M., Pape D., Harvey N., Schuck R., Ritter
E., Kohn S., Shinn T., Jackson T., Cardenas M., McCann R.,
Waterston R. and Wilson R.

Unpublished (1999)
The WashU-NCI Mouse EST Project 1999
Other ESTs: m85a07.x1
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63106, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LIND; contact the
IMAGE Consortium (info@image.lind.gov) for further information.
This read is a RESEQUENCE of a previously sequenced mouse clone
this read has been verified (found to hit its original self in the
correct orientation)

Seq primer: -40RP from gibco

[illegible]

ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 246)
AUTHORS	Dunn, D., Aoyagi, A., Barber, M., Becsorn, T., Duvall, B., Hamil, C., Ismail, R., Longacre, S., Malmoud, M., Meenen, E., Pedersen, T., Rolly, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A. and Wright, D., Weiss, R.
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL	Unpublished (2000)
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel.: 801 585 5606 Fax: 801 585 7177 Email: dduun@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0404 row: E column: 19 Seq primer: GCGTGTAAACGACGCCCACT Class: plasmid ends High quality sequence stop: 246. Location/Qualifiers 1..246
FEATURES	
Source	

FEATURES
source

BASE COUNT
ORIGIN

22 a 52 c 98 g 74 t

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0404E19"
/clone_lib="Mouse 10Kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: pMDA2ny; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMDA2 (g114732114[gb|AF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
```

[illegible]

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Anopheles.
1 (bases 1 to 1019)
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

2 (bases 1 to 1019)
Roth,C.W., Brey,P.T., Ke.Z., Collins,F.H. and Weissenbach,J.
Dirct Submission
Submitted (16-FEB-2000) BMJ, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

Location/Qualifiers
1 .1019
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="27H07"
/note="end : SP6"

BASE COUNT 120 a 286 c 296 g 304 t 13 others

ORIGIN

Query Match 18.7%; Score 61.6; DB 13; Length 1019;
Best local similarity 53.4%; Pred. No. 0.01;
Matches 124; Conservative 2; Mismatches 106; Indels 0; Gaps 0;

QY 85 caatgccagcccatgacaggcggtgaactgcagcagtccagtcggcccgggcagcy 144
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 955 GACAGCCCGGAGGCCACCACACTCTTCGGCAACCGTGC GGTCGAAGGACMACAGCACGCA 896
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 145 ggtcacgaatttttgatcgaccatgcggggcgcgctcaagtatgaccagcagcagcaga 204
/,
DB 895 GGAGCAGCGCAGCAGCAGCACATCAGCAGCGGGAACACAGCAGCAGCAGCAGCAGCA 836
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 205 gcacgacgacgacgacgacgacgacgaccttgcagccgcgcgcacccatccccaggaaac 264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dd 835 GCAGCAGCAGCAGCAGCAGCAGCAACAACACAGCAGCAGCGAACCAGCAGCGGTGAATG 776
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 265 cotcattaccaaaacctcgccaagtatcagcaactcaggcgagcagcagcgccag 316
/, /, /, /, /, /, /, /, /, /, /, /, /, /, /, /, /, /, /,
Dd 775 GCAGCAGCAGCAGCAGCAGCAGCATCAGCAGGAAACAGCAGCAGCAGCAGCAG 724
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AJ283665 LOCUS
DEFINITION 43B-AAB-B-12-F Anopheles gambiae immune competent 4A3B Anopheles
gambiae cDNA clone 4A3B-AAB-B-12, mRNA sequence.

ACCESSION AJ283665
VERSION
KEYWORDS EST...
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.

REFERENCE 1 (bases 1 to 452)
Dimopoulos,G., Casavant,T.L., Chang,S., Scheetz,T., Roberts,C.,
Donohue,M., Schultz,J., Benes,V., Borik,P., Ansorge,W., Soares,M.B.
and Kafatos,F.C.
TITLE Anopheles gambiae pilot gene discovery project: identification of
mosquito innate immunity genes from expressed sequence tags
generated from Immune-competent cell lines
Proc.Natl.Acad.Sci.U.S.A. 97 (12), 6619-6624 (2000)
20300950 MEDLINE

<hr/>					
RESULT	12				
CNS01GJA/C	DNA	GSS	01-JUN-2001		
LOCUS	1021 bp				
DEFINITION	Anopheles gambiae GSS T7 end of clone 06001 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.				
ACCESSION	AL143207				
VERSION	ALI43207.1	GI:7001325			
KEYWORDS	GSS.				
SOURCE	African malaria mosquito.				
ORGANISM	Anopheles gambiae				
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.				
REFERENCE	1 (bases 1 to 1021)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-FEB-2000) Genoscope - Centre National de Sequences : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
REFERENCE	2 (bases 1 to 1021)				
AUTHORS	Roth,C.W., Bray,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.				
TITLE	Direct Submission				
JOURNAL	Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France				
COMMENT	This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by Genoscope in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.				
FEATURES	Location/Qualifiers				
source	1. .1021				
	/organism="Anopheles gambiae"				
	/strain="PEST"				
	/db_xref="taxon:7165"				
	/clone="06001"				
	/clone_lib="NotreDamel"				
	/notes="end : T7"				
BASE COUNT	120 a 289 c 292 g 309 t 11 others				
ORIGIN					
	Query Match 18.9%; Score 62.4; DB 13; Length 1021;				
	Best Local Similarity 54.3%; Pred. No. 0.007;				
	Matches 126; Conservative 0; Mismatches 106; Indels 0; Gaps 0;				
QY	85 actgccccagcccatgaacaggccgtgaactgcagctccagctcgcccggcagcg 144 				
Db	955 GACAGCCGAGCGGCCACAGCTCTCGMAACCCKGCCTGGCCAGACCCACAGCAGA 896 				
QY	145 gtccagaattctatgctaccagtgcggcgccgtcagctatgccagcagcagca 204 				
Db	895 GGACAGCGGCGAGCAGCNACAGCATCACAGCGGGGAACAGCAGCAGCAGCAGCA 836 				
QY	205 gcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 264 				
Db	835 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 776 				
QY	265 ccttcattaccaaaactcgcaagtatcagcactacggggcagcaagccag 316 				
Db	775 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 724 				
<hr/>					
RESULT	13				
CNS01OJY/C	DNA	GSS	14-JUN-2001		
LOCUS	1019 bp				
DEFINITION	Anopheles gambiae GSS SP6 end of clone 27H07 of NotreDamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.				
ACCESSION	AL154154				
VERSION	ALI54154.1	GI:7015073			
KEYWORDS	GSS.				
SOURCE	African malaria mosquito.				
ORGANISM	Anopheles gambiae				

Probe #12713 for g
Probe #16768 used t
Probe #445 used t
Probe #448 for ge
Probe #604 used t
Probe #4437 used t
Probe #11714 for g
Probe #15750 used t
Probe #7457 used t
Probe #2526 for ge
Probe #2627 used t
Probe #2489 used t
Enucalyptus grandis
DNA sequence of hu
Receptor tyrosine
Human SCA7 genomic
Spheroecrallat at
Human gene express
DNA sequence of an
Human polynucleoti
M1891-containing p
Probe #13309 for g
Probe #17282 used t
Probe #8912 used t
Spheroecrallat at
An EcoRI fragment
Exons E, C and A o
Exons D, C, B and
Human adenosine A1
Polymorphic region
Glutamine rich reg
SCA2 gene CAG repe
Drosophila melanoga
Spheroecrallat at

[illegible][illegible]

[illegible]

Y 2050 tccgaggaccacatgagcagcttccctactgcaacacaggcccgctggcagccctgccc 2109
b 1420 tccgaggaccacatgagcagcttccctactgcaacacaggcccgctggcagccctgccc 1479
Y 2110 agggctcaacagcaactcgaagcgaaccccgagctcgtgtccaccccttctgtgacctct 2169
b 1480 agggctcaacagcaactcgaagcgaaccccgagctcgtgtccaccccttctgtgacctct 1539
Y 2170 cctgacgacatgtccacacaaatctgagcactcctccagagctccacaggaactgtccg 2229
b 1540 cctgacgacatgtccacacaaatctgagcactcctccagagctccacaggaactgtccg 1599
Y 2230 ctcgacagcttctccaaagctcgtgagcgggagcagctcgtccgctgctgctcagc 2289
b 1600 ctcgacagcttctccaaagctcgtgagcgggagcagctcgtccgctgctgctcagc 1659
Y 2290 gcoctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2349
b 1660 gcoctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1719
Y 2350 aagccgcgaacaaagctggcgtgaagcagcagcagcagcagcagcagcagcagcagcagc 2409
b 1720 aagccgcgaacaaagctggcgtgaagcagcagcagcagcagcagcagcagcagcagcagc 1779
Y 2410 ttctcgtggagcaacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 2469
b 1780 ttctcgtggagcaacacacagcagcagcagcagcagcagcagcagcagcagcagcagc 1839
Y 2470 cctggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2529
b 1840 cctggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1899
Y 2530 gactcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2589
b 1900 gactcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1959
Y 2590 aagcaaacctgtc 2649
b 1960 aagcaaacctgtc 2019
Y 2650 cccactacagcagctttgactgtttcccgagcaacacagcagcagcagcagcagcagcagc 2709
b 2020 cccactacagcagctttgactgtttcccgagcaacacagcagcagcagcagcagcagcagc 2079
Y 2710 gccaaaccccttgcctgcccagaggaacaaactggggatgttctccaggtggggattg 2769
b 2080 gccaaaccccttgcctgcccagaggaacaaactggggatgttctccaggtggggattg 2139
Y 2770 caccctggcagcttaccagggcctggagcagctggagagcagcagcagcagcagcagcagcagc 2829
b 2140 caccctggcagcttaccagggcctggagcagctggagagcagcagcagcagcagcagcagcagc 2199
Y 2830 aaaggggacacccatgagcttgcctgctgctgctgctgctgctgctgctgctgctgctgctg 2889
b 2200 aaaggggacacccatgagcttgcctgctgctgctgctgctgctgctgctgctgctgctgctg 2259
Y 2890 gagaagctgagcttgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2949
b 2260 gagaagctgagcttgcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2319
Y 2950 gaggcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 3009
b 2320 gaggcagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 2379
Y 3010 agccgagcagctgttccaccccgagcttgggagcagcagcagcagcagcagcagcagcagcagc 3069
b 2380 agccgagcagctgttccaccccgagcttgggagcagcagcagcagcagcagcagcagcagcagc 2439
Y 3070 aggaagc 3129
b 2440 aggaagc 2499
Y 3130 cccgagc 3189

RESULT 4
 AAD05635 standard; cDNA: 2409 BP.
 AAD05635;
 17-JUL-2001 (first entry)
 Human secreted protein-encoding gene 16 cDNA clone HDPY101, SEQ ID NO:67.
 Human; secreted protein; proliferative disorder; cancer; tumour;
 foetal abnormality; developmental abnormality; haematopoietic disorder;
 immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 inflammation; allergy; neurological disorder; Alzheimer's disease;
 Parkinson's disease; cognitive disorder; schizophrenia; asthma;
 skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
 cardiovascular disorder; angiotensin-related disorder; gene therapy;
 gastrointestinal disorder; pregnancy-related disorder; endocrine disorder;
 endocrine disorder; infection; wound healing; vulnery;
 cell culture; chemotaxis; food additive;
 binding partner identification; ss.
 Homo sapiens.
 Key Location/Qualifiers
 CDS 33..329
 /tag= a
 /product= "Human secreted protein"
 sig_peptide 33..143
 /tag= b
 144..326
 /tag= c
 /product= "Human mature secreted protein"
 mat_peptide 144..326
 /product= "Human mature secreted protein"
 WO200134627-A1.
 17-MAY-2001.
 08-NOV-2000; 2000MO-US30628.
 12-NOV-1999; 99US-0164744.
 30-JUN-2000; 2000US-0215140.
 (HUMA-) HUMAN GENOME SCI INC.
 Ruben SM, Komatsoulis GA, Baker KP, Young PE;
 WPI: 2001-316481/33.
 P-PSDB; AA001826.
 New nucleic acid molecules encoding human secreted proteins, used in
 preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
 Parkinson's diseases and cancers -
 Claim 1; Page 465; 567pp; English.
 AAD05579-AAD05658 represent cDNAs corresponding to 28 human secreted
 protein genes and AAD01770-AA001849 represent the proteins they encode.
 AAD01850-AA001860 represent human secreted protein fragments or variants.
 The genes and their secreted proteins are useful for preventing,
 treating or ameliorating medical conditions, e.g., by protein or gene
 therapy. Pathological conditions can be diagnosed by determining the
 amount of the new protein in a sample or by determining the presence of
 mutations in the new genes. Specific uses are described for each of the
 28 genes, based on the tissues in which they are most highly expressed,
 and include developing products for the diagnosis or treatment of
 proliferative disorders, cancer, tumours, foetal and developmental
 abnormalities, haematopoietic disorders, diseases of the immune system,
 AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
 allergies, neurological disorders (e.g., Alzheimer's disease,
 Parkinson's disease), cognitive disorders, schizophrenia, asthma,
 skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
 cardiovascular disorders, angiotensin-related disorders, kidney disorders,

gastrointestinal disorders, pregnancy-related disorders, endocrine
 disorders, and infectious. The proteins can also be used to aid wound
 healing and epithelial cell proliferation, to prevent skin aging due to
 sunburn, to maintain organs before transplantation, for supporting cell
 culture of primary tissues, to regenerate tissues, to identify their
 cognate ligands or binding partners, and in chemotaxis, and can be used
 as a food additive or preservative to modify storage properties.
 Antibodies specific for a protein of the invention can be used in
 alleviating symptoms associated with the disorders mentioned above, and
 in diagnostic immunoassays (e.g., radioimmunoassay or enzyme linked
 immunosorbent assay (ELISA)). The present sequence represents a human
 secreted protein-encoding cDNA of the invention.

Sequence 2409 BP; 490 A; 751 C; 711 G; 450 T; 7 other;

Query Match 11.9%; Score 714; DB 22; Length 2409;
 Best Local Similarity 99.5%; Pred. No. 3.9e-128;
 Matches 726; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

5295 gcgagagcgcttaccacacatgacatgtgtcaactccctcgagagagcgcccaagc 5354
 |||||
 12 gcgagagcgcttaccacacatgacatgtgtcaactccctcgagagagcgcccaagc 71
 5355 ccacaggaagccttctgga 5414
 |||||
 72 ccacaggaagccttctgga 131
 5415 tgcacgagcgcttctgga 5474
 |||||
 132 tgcacgagcgcttctgga 191
 5475 ctgacgagcgcttctgga 5534
 |||||
 192 ctgacgagcgcttctgga 251
 5535 ttgctctttgt 5594
 |||||
 252 ttgctctttgt 311
 5595 tgggacgagcgcttctgga 5654
 |||||
 312 tgggacgagcgcttctgga 371
 5655 ggcgacgagcgcttctgga 5714
 |||||
 372 ggcgacgagcgcttctgga 431
 5715 cgaagtgtgagcttctgga 5772
 |||||
 432 cgaagtgtgagcttctgga 491
 5773 ggcgacgagcgcttctgga 5832
 |||||
 492 ggcgacgagcgcttctgga 551
 5833 ctactgtctgt 5892
 |||||
 552 ctactgtctgt 611
 5893 caaacatgagcttctgga 5952
 |||||
 612 caaacatgagcttctgga 671
 5953 ggtacagagcgcttctgga 6012
 |||||
 672 ggtacagagcgcttctgga 731
 6013 tgggctgagc 6022
 |||||
 732 tgggctgagc 741

DB '238 ccaagcaccatcaccc 250

RESULT 7
ID AAH3839/c
ID AAH3839 standard; cDNA; 158 BP.
AC AAH3839;
XX
XX
XX 03-SEP-2001 (first entry)
XX
XX
XX Human colon cancer antigen encoding cDNA SEQ ID NO:895.
DE
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KM colorectal carcinoma; ss.
XX
XX Homo sapiens.
OS
XX
XX W0200122920-A2.
PN
XX
XX 05-APR-2001.
PD
XX
XX 28-SEP-2000; 2000MO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX
XX 03-NOV-1999; 99US-0163280.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA.
PI
XX
XX WPI: 2001-235357/24.
DR
XX
XX P-PSDB; AAG74408.
PT
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX
XX Claim 1; Page 2817; 9803pp; English.

AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing P, inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated PS, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention.

Fig. 6: Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.

Sequence 158 BP; 20 A; 45 C; 63 G; 30 T; 0 other;

Query Match 2.6%; Score 158; DB 22; Length 158;
Best Local Similarity 100.0%; Pred. No. 1,5e-21;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 3552 atgtgtactcgtctctctcagagcccttgatgagcccgccgagccgagccccaagcgtc 3711
DB 158 ATGTGACTCGTCTCTCTCAGAGCCCTTGATGAGCCCGCCGAGCCCGAGCCCGAGCCGCTG 99
QY 3712 acacacacccttcacccacagacaacctggggggaagcagcagcagcgcgcttcaagtcg 3771
DB 98 ACCACACCCCTTCACCCCGAGACAACCTGGGGGGGCAAGGAGAGAGCGCGCTTCAAGTCG 39

RESULT 10

AAQ20685 standard; DNA; 14704 BP.

AAQ20685;
22-APR-1992 (first entry)
PKS 741 insert contg. SRY.

Y-chromosome; SRY; sex-determining region; genital ridge; ss.

Mus musculus.

M09200375-A.

09-JAN-1992.

28-JUN-1991; 91MO-GB01057.

09-MAY-1991; 91GB-0010085.

28-JUN-1990; 90GB-0014446.

13-JUL-1990; 90GB-0015488.

(IMCR) IMP CANCER RES TECH.

Goodfellow FW, Lovell-Badge R;

WPI; 1992-041557/05.

Nucleic acid pY5.3, its fragments and oligo-nucleotide(s) - is
used for determining or controlling sex of e.g. embryos by
detecting Y chromosomes

Claim 7; Page 122 and Fig 19(1-3); 183pp; English.

The mouse gene equivalent to the human Y-located gene defined by
pY5.3 (AAQ20684), is referred to as sex-determining Region of the
Y-chromosome (SRY) (formerly (mating type box Y) MY).

The mouse SRY is expressed in male genital ridge, at the time that
sex determination occurs, and adult testis.

See also AAQ20680-81, AAQ20684-87, AAQ2530-31, AAQ2535-36 and
AAQ2560-67.

Sequence 14704 BP; 4259 A; 3237 C; 3029 G; 4103 T; 6 other;

Query Match 1.4%; Score 85.4; DB 13; Length 14704;

Best Local Similarity 46.4%; Pred. No. 3, 1e-07;
Matches 278; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 905 ccaagatgatgaggaactgatgaaagagcagcgtgccccagcagcagcgtatgag 964
DB 8706 ccaaaacacagcaagctttatctgagcagcgtgatatcccatctggcactgagc 8765
QY 965 agcagggcgcagcagcgtgccccttggaactactccctgacagcagccagccgc 1024
DB 8766 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8825
QY 1025 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1084
DB 8826 tctatgaccacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8885
QY 1085 cccctctgcttcccccagcagcagcagcagcagcagcagcagcagcagcagc 1144
DB 8886 accaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8945
QY 1145 cctcaccactact 1204
DB 8946 accaccaccacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 9005

RESULT 11

AAQ3007 standard; DNA; 10266 BP.

AAQ3007;

23-OCT-1996 (first entry)

Mouse SRY-related gene.

Mouse; SRY; primer; PCR; polymerase chain reaction; amplification; probe;
HMG box; human; bovine; sex; animal; birth; ds.

Mus musculus.

Key

CDS

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

QY 1205 gcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1264
DB 9006 atgacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 9065

QY 1265 cggggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1324
DB 9066 agaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 9125

QY 1325 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1384
DB 9126 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 9185

QY 1385 cccaggaacaccttcattcaacaaccttcgcaagcagcagcagcagcagcagcagc 1444
DB 9186 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 9245

QY 1445 agggctactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1503
DB 9246 accaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 9304

Location/Qualifiers
7148..8335
/tag= a
/product= SRY-related protein

JP08154685-A.
18-JUN-1996.
30-NOV-1994; 94UP-0319525.
30-NOV-1994; 94UP-0319525.

(KACH-) KACHIKU JUSEIRAN ISHOKU GIKOTSU KENKYUKU.
WPI; 1996-336575/34.
P-PSDB; AAQ03474.

Bovine and mouse SRY-related DNA - useful for detecting e.g. the sex
of unborn animals
Claim 1; Page 10-14; 21pp; Japanese.

This is the nucleotide sequence of a gene encoding a mouse SRY-related
protein. The gene was isolated from a mouse genomic library using a
cDNA fragment amplified by primers AAT3009-10 as a probe. The screen
isolated 4 EORI fragments of 2.3, 2.8, 3.5 and 1.5 kb covering the
gene. Sequence analysis revealed a 240 bp HMG box sequence between bases
7154-7393. Similarity with the human SRY HMG box sequence resulted in
primers being generated to amplify the human SRY HMG box sequence for
use as a probe to isolate the bovine SRY-related gene (AAT33008). The
mouse and bovine genes are useful for determining the sex of an animal
prior to birth.

Sequence 10266 BP; 2958 A; 2202 C; 2228 G; 2678 T; 0 other;


```
Query Match      1.48; Score 82.2; DB 17; Length 10266;
Best Local Similarity 46.1%; Pred. No. 1.2e-06;
Matches 276; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

y 905 ccaagtatgatgagaactgtatgaataaacacagcagtcgcccccacgaggcagcgacgtatgcag 964
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7521 caaaataaccggaaaagctttattggcagctgttgatatccccactgggcacctgcagc 7580
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 965 agcagggcgccacagtgccttttgcgactcacctccctgcacgtccagtcacgcccacgcgc 1024
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7581 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7640
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1025 ccagcagccctcggcataccocagctccaaggcggaagctgcagaagcagcattgcct 1084
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7641 tctatgacacaccacgacgacgacgacgacgacgacgacgacgacgacgacgttccatgcac 7700
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1085 ccctctgccttccccagggtaccacacttcttcacgacttcacgtcttcccacact 1144
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7701 accaccagggaagcaggaggtttcatgaccaccaaccagcagcaacagcagttccatgacc 7760
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1145 ctcccacttactctctctgtccagggtggtgggcaggggggccacactcctctaagagtt 1204
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7761 accaccaccaaccaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7820
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1205 gcacagcacccagctgccagagcccatatcacggccctgactgccagcctcagcctggcc 1264
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7821 atgaccaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7880
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1265 cgggcgagcgggtccgaatcttcctgcttacagctcgggcgccctcagctatgaccagc 1324
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7881 agaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 7940
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1325 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 7941 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8000
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1385 ccaggaacaacctccaataccaaaactccgaagtatccagcactcagcagggcagcaagcc 1444
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 8001 ccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8060
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
y 1445 agggctactgcgacgcgggagcagcgcgtccggagcccccagcagcagcactaccacagccttc 1503
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 8061 accaccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 8119
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 12	
AAAI22780	
ID	AAI22780 standard; DNA; 1635 BP.
AC	
XX	AAI22780;
XX	
XX	12-OCT-2001 (first entry)
XX	
XX	Probe #12713 for gene expression analysis in human cervical cell sample.
DE	
XX	Probe; human; microarray; gene expression; cervical epithelial cell;
KK	cervical cancer; ss.
KK	
XX	
XX	Homo sapiens.
OS	
XX	WO200157278-A2.
PN	
XX	
PD	09-AUG-2001.
PD	
PF	30-JAN-2001; 2001WO-US00670.
PF	
XX	
XX	04-FEB-2000; 2000US-0180312.
XX	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.

[illegible]

b 382 gccgcgcgc 389

RESULT 3
 US-08-232-463-14
 ; Sequence 14, Application US/08232463
 ; Patent No. 5670367
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: DORNER, F.
 ; APPLICANT: SCHEIFFLINGER, F.
 ; APPLICANT: FALKNER, F. G.
 ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 ;
 ; NUMBER OF SEQUENCES: 52
 ;
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: FOLEY & Lardner
 ; STREET: 1800 Diagonal Road, Suite 500
 ; CITY: Alexandria
 ; STATE: VA
 ;
 ; COUNTRY: USA
 ;
 ; ZIP: 22313-0299
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ;
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/232.463
 ;
 ; FILING DATE:
 ;
 ; CLASSIFICATION: 435
 ;
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/935.313
 ;
 ; FILING DATE:
 ;
 ; APPLICATION NUMBER: EP 91 114 300.6

RESULT 144
US-08-209-747-1/c
Sequence 1, Application US/08209747
Patent No. 5733771
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Coldin, Mark
TITLE OF INVENTION: CDNAs Encoding Minor Ampullate Spider
TITLE OF INVENTION: Silk Proteins
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSES: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/209, 747

FILING DATE: 14-MAR-1994
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1447-104P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-205-8000
 TELEFAX: 703-205-8050
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2793 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Nephila clavipes
 TISSUE TYPE: minor ampullate gland
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 183..2675
 OTHER INFORMATION: /product="N. clavipes minor
 OTHER INFORMATION: ampullate silk protein"
 US-08-209-747-1

Query Match 1.18; Score 63.6; DB 1; Length 2793;
 Best Local Similarity 51.0%; Pred. No. 0.0005;
 Matches 203; Conservative 0; Mismatches 189; Indels 6; Gaps 2;

QY 1269 ggaagcggtcccaaatcttctatgctaccagtcggcgccgtacgtatggccagcagga 1328
 DB 1638 gcaactgcccacacacgacgacgacgacgacgacgacgacgacgacgacgacgac 1579
 QY 1329 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1388
 DB 1578 cctcgagctcctgtgtgacagcagcagcagcagcagcagcagcagcagcagcagcag 1519
 QY 1389 ggaacccctccattaccaaaacctgcagcaagatagactagcggcagcagcagcagc 1448
 DB 1518 ccttgaacacgtagacgcttcacgctcctgacgctcctgacgctcctgacgctcctgac 1459
 QY 1449 ctactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1508
 DB 1458 ccttgaacacgtagacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1404
 QY 1509 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1568
 DB 1403 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1345
 QY 1569 gtcgcagctgtagcacaacctggagaaacttccctacagcagcagcagcagcagcagc 1628
 DB 1344 gaaagagctcctctctcagacacccgacgagcagcagcagcagcagcagcagcagc 1285
 QY 1629 ggccttcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1666
 DB 1284 AACCAAGAGGCTCCAGCTCCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1247

RESULT 5
 US-08-458-298-1/c
 Sequence 1, Application US/08458298
 Patent No. 5758677
 GENERAL INFORMATION:
 APPLICANT: Lewis, Randolph V.
 APPLICANT: Colgild, Mark
 TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
 TITLE OF INVENTION: Silk Proteins
 NUMBER OF SEQUENCES: 56
 CORRESPONDENCE ADDRESS:
 ADDRESS: Birch, Stewart, Kolasch & Birch

STREET: P.O. Box 747
 CITY: Falls Church
 STATE: Virginia
 COUNTRY: USA
 ZIP: 22040-3487
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/458,298
 FILING DATE: 02-JUN-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/209,747
 FILING DATE: 14-MAR-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 1447-104P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-205-8000
 TELEFAX: 703-205-8050
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2793 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: CDNA
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Nephila clavipes
 TISSUE TYPE: minor ampullate gland
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 183..2675
 OTHER INFORMATION: /product="N. clavipes minor
 OTHER INFORMATION: ampullate silk protein"
 US-08-458-298-1

Query Match 1.18; Score 63.6; DB 1; Length 2793;
 Best Local Similarity 51.0%; Pred. No. 0.0005;
 Matches 203; Conservative 0; Mismatches 189; Indels 6; Gaps 2;

QY 1269 ggaagcggtcccaaatcttctatgctaccagtcggcgccgtacgtatggccagcagga 1328
 DB 1638 gcaactgcccacacacgacgacgacgacgacgacgacgacgacgacgacgacgac 1579
 QY 1329 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1388
 DB 1578 cctcgagctcctgtgtgacagcagcagcagcagcagcagcagcagcagcagcagcag 1519
 QY 1389 ggaacccctccattaccaaaacctgcagcaagatagactagcggcagcagcagcagc 1448
 DB 1518 ccttgaacacgtagacgcttcacgctcctgacgctcctgacgctcctgacgctcctgac 1459
 QY 1449 ctactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1508
 DB 1458 ccttgaacacgtagacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 1404
 QY 1509 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1568
 DB 1403 cagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1345
 QY 1569 gtcgcagctgtagcacaacctggagaaacttccctacagcagcagcagcagcagcagc 1628
 DB 1344 gaaagagctcctctcagacacccgacgagcagcagcagcagcagcagcagcagc 1285
 QY 1629 ggccttcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1666

Db 1284 AACACAGCGCTCCAGCTCTGCTCCAGCAGCTGCA 1247

RESULT 6

US-08-469-802B-3
Sequence 3, Application US/08469802B

Patent No. 5741645
GENERAL INFORMATION:

APPLICANT: Orr, Harry T.
APPLICANT: Rannum, Laura P.W.
APPLICANT: Chung, Ming-yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

Patent No. 5741645
TITLE OF INVENTION: Type 1 and Method for Diagnosis

NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Muehting, Raasch, Gebhardt & Schwappach, P.A.
STREET: 119 No. 5741645th Fourth Street, Suite 203
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,802B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Muehting, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 110.00030101

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1225

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 234 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-469-802B-3

Query Match 1.0%; Score 61.4; DB 1; Length 234;

Best Local Similarity 54.7%; Pred. No. 0.00072;

Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323

Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66

QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383

Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126

QY 1384 gccaggaacacctccattccaaacctgccagctacgactacggtcgagcaaggc 1443

Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186

QY 1444 cagggctactgcagccgagcagcagcagcagcagcagcagcagcagcagc 1486

Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

RESULT 7

US-08-267-803B-3

Sequence 3, Application US/08267803B

Query Match 1.0%; Score 61.4; DB 1; Length 234;

Best Local Similarity 54.7%; Pred. No. 0.00072;

Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323

Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66

QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383

Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126

QY 1384 gccaggaacacctccattccaaacctgccagctacgactacggtcgagcaaggc 1443

Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186

QY 1444 cagggctactgcagccgagcagcagcagcagcagcagcagcagcagcagc 1486

Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

Patent No. 5834183

GENERAL INFORMATION:

APPLICANT: Orr, Harry T.
APPLICANT: Rannum, Laura P.W.
APPLICANT: Chung, Ming-yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia

Patent No. 5834183

TITLE OF INVENTION: Type 1 and Method for Diagnosis

NUMBER OF SEQUENCES: 85

CORRESPONDENCE ADDRESS:

ADDRESSEE: Muehting, Raasch, Gebhardt & Schwappach, P.A.

STREET: P.O. Box 581415

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55458-1415

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/267,803B

FILING DATE: 28-JUN-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: McCormack, Myra H.

REGISTRATION NUMBER: 36,602

REFERENCE/DOCKET NUMBER: 110.00030120

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-305-1217

TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 234 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

US-08-267-803B-3

Query Match 1.0%; Score 61.4; DB 2; Length 234;

Best Local Similarity 54.7%; Pred. No. 0.00072;

Matches 122; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 1264 ccggggcagcgggtccagaattcttcagtcggtcgccgctcagctatgaccag 1323

Db 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66

QY 1324 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383

Db 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126

QY 1384 gccaggaacacctccattccaaacctgccagctacgactacggtcgagcaaggc 1443

Db 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186

QY 1444 cagggctactgcagccgagcagcagcagcagcagcagcagcagcagcagc 1486

Db 187 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 229

RESULT 8

US-08-232-463-14/C

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEIFLINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232,463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935,313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)836-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: pTZgpt-F1s
 US-08-232-463-14

Query Match 1.0%; Score 61.4; DB 1; Length 7218;
 Best Local Similarity 1.3%; Pred. No. 0.0018;
 Matches 5; Conservative 233; Mismatches 139; Indels 0; Gaps 0;

```

QY 11 ggcccaaggagatgagagcgaggaattgctgcttaattctttagcttcaggaag 70
DB 1439 GgTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 71 ggcctgagatgagtgagtgaggaattgaggaagcgctgcaatggcgagggagcg 130
DB 1379 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 131 ggtagctccacgtgacgtgagcgctgagcgagcggtttgctccgcctagtgccgt 190
DB 1319 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 191 ccttggcaagctcagctgagcgatgttgagcgaggaattacagagacacccctct 250
DB 1259 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 251 aacacgaagtgtcatatagcgacacgctgttgaggaagctcagggaggggaactat 310
DB 1199 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 311 taagagctgcgagggagcaggaaggttgaggaaggtggttgaggaaggtctcttagg 370
DB 1139 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
QY 371 caaaggaagtgcgcc 387
DB 1079 RRRRRRRRRRRRRRATCG 1063
  
```

RESULT 9
 US-09-043-303-7
 Sequence 7, Application US/09043303
 Patent No. 6251589
 GENERAL INFORMATION:
 APPLICANT: TSUJI, Shoji
 TITLE OF INVENTION: Method for Diagnosing SpinoCerebellar Ataxia Type 2 and
 TITLE OF INVENTION: Primers therefor
 FILE REFERENCE: 0760-0241P
 CURRENT APPLICATION NUMBER: US/09/043,303
 CURRENT FILING DATE: 1998-05-18
 EARLIER APPLICATION NUMBER: PCT/JP96/01999
 EARLIER FILING DATE: 1996-07-18
 NUMBER OF SEQ ID NOS: 17
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 7
 LENGTH: 203
 TYPE: DNA
 ORGANISM: P-2093 plasmid
 US-09-043-303-7

Query Match 1.0%; Score 58.8; DB 4; Length 203;
 Best Local Similarity 56.1%; Pred. No. 0.0023;
 Matches 111; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

```

QY 1263 ccgggggacaggggtcaggaattctatgctaccagtcgggcctcagctatgacca 1322
DB 6 ccagcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagca 65
QY 1323 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagca 1382
DB 66 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagca 125
QY 1383 tgcacaggaacccctcatltaacaaacctgcgaagatcacagcactacggaggaag 1442
DB 126 gtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagca 185
QY 1443 ccaggctactcagcc 1460
DB 186 tcacggaactctggcc 203
  
```

RESULT 10
 US-08-965-903B-1
 Sequence 1, Application US/08965903B
 Patent No. 6060275
 GENERAL INFORMATION:
 APPLICANT: Hachon, NLR
 APPLICANT: Krasnow, Mark A.
 TITLE OF INVENTION: SPROUTY PROTEIN AND CODING
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 STREET: 350 Cambridge Ave., Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/965,903B
 FILING DATE: 07-NOV-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/030232

FILING DATE: 07-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Petithory, Joanne R.
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 8600-0177.30
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3254 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE: Drosophila spry cdna
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1323...3095
OTHER INFORMATION:
US-08-965-903B-1

Query Match 1.0%; Score 58.2; DB 3; Length 3254;
Best Local Similarity 59.3%; Pred. No. 0.0062;
Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1213 ccgaactgcccagcccatgacagcgcctgacctgcaagtcggcgcccttgcagagccggcgag 1272
DB 1365 CCGGCTGCCAGCAGGAGTAATGGCCAGACGACCACCACTGTTCGTGCCCCGC 2024
QY 1273 cgggttcgaattttatgacctaccagtgcggcgcccttgatgaccagcagcagcag 1332
DB 2025 CGGAATFCAGCACTTCACTTGCAACAACACGACGACGAGCATCTGCAGCACGAG 2084
QY 1333 cagcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1379
DB 2085 CAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2131

RESULT 11
US-08-469-802B-2
Sequence 2, Application US/08469802B
Patent No. 5741645
GENERAL INFORMATION:
APPLICANT: Orr, Harry T.
APPLICANT: Ranum, Laura P.W.
APPLICANT: Chung, Ming-yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
PATENT NO. 5741645
TITLE OF INVENTION: Type 1 and Method for Diagnosis
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
STREET: 119 N. 5741645th Fourth Street, Suite 203
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,802B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mueting, Ann M.

FILING DATE: 07-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Petithory, Joanne R.
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 8600-0177.30
TELEPHONE: 650-324-0880
TELEFAX: 650-324-0960
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3254 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE: Drosophila spry cdna
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1323...3095
OTHER INFORMATION:
US-08-965-903B-1

Query Match 1.0%; Score 58.2; DB 3; Length 3254;
Best Local Similarity 59.3%; Pred. No. 0.0062;
Matches 99; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 1213 ccgaactgcccagcccatgacagcgcctgacctgcaagtcggcgcccttgcagagccggcgag 1272
DB 1365 CCGGCTGCCAGCAGGAGTAATGGCCAGACGACCACCACTGTTCGTGCCCCGC 2024
QY 1273 cgggttcgaattttatgacctaccagtgcggcgcccttgatgaccagcagcagcag 1332
DB 2025 CGGAATFCAGCACTTCACTTGCAACAACACGACGACGAGCATCTGCAGCACGAG 2084
QY 1333 cagcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1379
DB 2085 CAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2131

RESULT 11
US-08-469-802B-2
Sequence 2, Application US/08469802B
Patent No. 5741645
GENERAL INFORMATION:
APPLICANT: Orr, Harry T.
APPLICANT: Ranum, Laura P.W.
APPLICANT: Chung, Ming-yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
PATENT NO. 5741645
TITLE OF INVENTION: Type 1 and Method for Diagnosis
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
STREET: 119 N. 5741645th Fourth Street, Suite 203
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,802B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mueting, Ann M.

[illegible][illegible]

Search completed: December 4, 2001, 16:11:33
Job time: 7056 sec

Query Match	9.6%;	Score 575.6;	DB 11;	Length 1012;
Best Local Similarity	95.0%;	Pred. No. 1.1e-90;		
Matches 632;	Conservative	0;	Mismatches 21;	Indels 5;
				Gaps 4;

Mr's
Am Klopferstift 18a D-82155 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de,
sequenced by Medicinomix (Martinsried/Germany) within the CDNA
sequencing consortium of the German Genome Project. No 51 sequence
available.
This clone (DKFZ761A121) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Query Match	8.8%;	Score 527.2;	DB 10;	Length 566;
Best Local Similarity	98.9%;	Pred. No. 2.7e-82;		

Contract: matia.washington@medline.wustl.edu
Washington University School of Medicine
4444 Forest Park Parkway, Box 9501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouse@wustl.wustl.edu
This clone is available royalty-free through INTL; contact the
IMAGe Consortium (info@imge.intl.gov) for further information

TITLE	JOURNAL MEDLINE COMMENT
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 20202663
Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01505-010, Sao Paulo-SP, Brazil	

Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@uic.edu or br.felipe@uic.edu
 This sequence was derived from the PAPSP/HCC Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.uic.edu/org.br/scripts/gethtml2.pl?L=5c2-RC0-C10379-066>)
 200-032-b086t3-2000-02-06s4-1)
 Seq primer: puc 18 forward
 High quality sequence start: 38
 High quality sequence stop: 558.

L. 358
Jordaniana "Homo sapiens"

BASE COUNT	98 a	207 c	147 g	106 t
ORIGIN				

Best Local Similarity 98.68; Pred. No. 6.5e-80;
Matches 550; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

Db : 2 ACCCTCCTGCGGGTGAG-AGGAGACGGCTCACCAACCATAATGCACTGTTGCACTC 60

Db 61 CCTGTGAGATGCGGCCCAAGCCCCACACAGGAAGCCTTCTCCTCTGCTCCTCTTCTCAT 120

db 121 CCGCGCTCTCTCTCCCTGGATGCAGCCGGGGCCCTCCCTGGCCACACTCCCTGGAGGCT 180

Db 181 CCATCCTGCAAGCGCGGCCCTTCCTTGCCTCCCTCTCTCCACGATACACATGGGCCCTGG 240

5573 tcaagaccttgggacctctgttggccctaactaacctgaacactgctctcccaaaaga 563

5633 agccaaactcaagagagaggtgctgcagagaagcacctgtgagagagcctcgtgcgcg 569

QY 5693 ttgagagaaacactcaaggtcccgagttgtgcagctgcgcgcactgcgcggaagccccc 575

/dev_stage="Adult"
 /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 97 a 203 c 140 g 100 t
 ORIGIN

Query Match 8.4%; Score 506; DB 10; Length 540;
 Best Local Similarity 98.5%; Pred. No. 1.4e-78;
 Matches 532; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 5291 aagaagcagcagcgttcaacacatgatgacgtgttaactcccttgatgagtcgccc 5350
 DB 2 AGAAGGAGAGCGG-TCACACACATATGCACTGTGTGTCACCTCCCTGAGATGCGGCC 60
 QY 5351 agcccccaggaagccttccctctctctctctctctctctctctctctctctct 5410
 DB 61 AGCCCCAGAGAGCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 120
 QY 5411 tagatgcagcggggtcctccttgccacacactcccttgagggctcctcctcctcctcct 5470
 DB 121 TGGATGACGCGGGGCTTCCTGCGCACACTCCCTGAGGCTCATCTGACGCCGCGC 180
 QY 5471 cctccttgccctctctctctcctcctcctcctcctcctcctcctcctcctcctcct 5530
 DB 181 CCTCCTGCCCCCT 240
 QY 5531 cctcttgctgttttctgctctctctcctcctcctcctcctcctcctcctcctcctcct 5590
 DB 241 CCTCTTGCTTTGTTTCT 300
 QY 5591 tctgtggcctctactacccctgaacactcctcctcctcctcctcctcctcctcctcct 5650
 DB 301 TCTGTGGCCCTACTACCTGTAACATGCTGCTCCCAAAAAGCAAACTCAAGAGAG 360
 QY 5651 aggtgcggcagaagagcactctgtgagagagcctcctcctcctcctcctcctcctcct 5710
 DB 361 AGGTGCGGCCAGAGGACCTGTGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
 QY 5711 gtccgagctgtgagctgcgcacactgcgggaagccccccag--tgagccgcaactg 5768
 DB 421 GTCCGAGTGTGAGCTGCGGACACTGCGGAGAGCCCCAGCCTGACGSCCCTACTG 480
 QY 5769 accgcgcaagcggcgcacactgcgcacagtgcgcggggcctcctcctcctcctcctcct 5828
 DB 481 ATCCGCGCCAGAGGCGCCACTGCGCACAGTGCCTGCGGCTGTCCGAGGCTGCAGA 540

RESULT 12

LOCUS AM821899 540 bp mRNA EST 17-MAY-2000
 DEFINITION RC0-ST0379-210100-032-d02 ST0379 Homo sapiens cDNA, mRNA sequence.
 VERSION AM821899
 KEYWORDS AM821899.1 GI:7914893
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 540)
 Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, N. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brunstein, A., de Oliveira, P.S., Bucher, P., Tongenel, C.V., O'Hare,
 M.J., Soares, F., Brentani, K.R., Reis, D.F., de Souza, S.J., and
 Simpson, A.V.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LIC Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?L=412-RC0-ST0379-210>)
 100-032-d02&t3=2000-01-21&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 20
 High quality sequence stop: 540.
 Location/Qualifiers

FEATURES

1..540
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="ST0379"
 /dev_stage="Adult"
 /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT 97 a 203 c 140 g 100 t
 ORIGIN

Query Match 8.4%; Score 506; DB 10; Length 540;
 Best Local Similarity 98.5%; Pred. No. 1.4e-78;
 Matches 532; Conservative 0; Mismatches 5; Indels 3; Gaps 2;

QY 5291 aagaagcagcagcgttcaacacatgatgacgtgttaactcccttgatgagtcgccc 5350
 DB 2 AGAAGGAGAGCGG-TCACACACATATGCACTGTGTGTCACCTCCCTGAGATGCGGCC 60
 QY 5351 agcccccaggaagccttccctctctctctctctctctctctctctctctctct 5410
 DB 61 AGCCCCAGAGAGCCTTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 120
 QY 5411 tagatgcagcggggtcctccttgccacacactcccttgagggctcctcctcctcctcct 5470
 DB 121 TGGATGACGCGGGGCTTCCTGCGCACACTCCCTGAGGCTCATCTGACGCCGCGC 180
 QY 5471 cctccttgccctctctctcctcctcctcctcctcctcctcctcctcctcctcctcct 5530
 DB 181 CCTCCTGCCCCCT 240
 QY 5531 cctcttgctgttttctgctctctctcctcctcctcctcctcctcctcctcctcctcct 5590
 DB 241 CCTCTTGCTTTGTTTCT 300
 QY 5591 tctgtggcctctactacccctgaacactcctcctcctcctcctcctcctcctcctcctcct 5650
 DB 301 TCTGTGGCCCTACTACCTGTAACATGCTGCTCCCAAAAAGCAAACTCAAGAGAG 360
 QY 5651 aggtgcggcagaagagcactctgtgagagagcctcctcctcctcctcctcctcctcct 5710
 DB 361 AGGTGCGGCCAGAGGACCTGTGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
 QY 5711 gtccgagctgtgagctgcgcacactgcgggaagccccccag--tgagccgcaactg 5768
 DB 421 GTCCGAGTGTGAGCTGCGGACACTGCGGAGAGCCCCAGCCTGACGSCCCTACTG 480
 QY 5769 accgcgcaagcggcgcacactgcgcacagtgcgcggggcctcctcctcctcctcctcct 5828
 DB 481 ATCCGCGCCAGAGGCGCCACTGCGCACAGTGCCTGCGGCTGTCCGAGGCTGCAGA 540

QY : 2255 cqqqtgaqcqqqactatccq 2274

Search completed: December 4, 2001, 14:52:51
Job time: 6785 sec

||||| 890 TACCTGAGGGTGGCCAGTAGAGAACTTGAGAAAGACAGCA 939
||||| 151 ValProProSerArgIntyrAlaGluGlnValProPheAr 167
||||| 940 GTCCGCCCCAGCAGCAGTAGATGACAGCAGGCGCCAGGTGCCCTTGG 989
167 gThrHisSerLeuHisValGlnGlnProProProGlnGlnProLeuA 184
990 GATCATCTCCCTCCAGTCCAGCAGCCAGCCGCCGCCGCCAGCAGCCCTGG 1039
184 latyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerPro 200
1040 CATACCCCAAGCTCCAAAGGACAGAGCTGCAGAACACATTGCCCTCCCT 1089
201 LeuProPheProGlnGlnThrHisPheProGlnHisSerGlnSerPhePr 217
1090 CTGCCCTTCCGCCAGGTACCCACTTCTCCAGCATTCCTCCAGTCCCTCCC 1139
217 oThrSerSerThrTyrSerSerValGlnGlnGlyGlyGlnGlyAlaH 234
1140 CACTCTCCACCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1189
234 isSerTyrLysCysThrAlaProThrAlaGlnProHisAspArgPro 250
1190 ACTCTATTAAGAGTTGCACAGCAGCCAGCTGCCAGCCCTATGACAGCCG 1239
251 LeuThrAlaSerSerSerLeuAlaProGlnArgValGlnAsnLeuHi 267
1240 CTGACTGCCAGCTCCAGCTTGGCCCGGGGCGAGGGCTCCAGATCTCA 1289
287 salATyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGln 284
1290 TGCTTACCATGCGGGCGCTCAGTATGACAGCAGCAGCAGCAGCAGC 1339
284 lngGlnGlnGlnGlnGlnGlnAlaLeuGlnSerArgHisAlaGln 300
1340 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1389
301 GluThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlyGln 317
1390 GAACCTCTCATTAACAAACCTCCGCCAGTATCAGCATCAGGCGCAGCA 1439
317 nGlyGlnGlyTyrCysGlnProAspAlaAlaValArgThrProGlnGln 334
1440 AGGCCAGGCTACTGCCAGCCGAGCAGCAGCAGCAGCAGCAGCAGCAGT 1489
334 yrTyrGlnPhePheSerProSerSerHisSerProAlaArgSerVal 350
1490 ACTACAGACCTTCAGCCCGCAGCTCCAGCCACTCCGCCCGCGCTCCGTG 1539
351 GlyArgSerProSerTyrSerThrProSerProLeuMetProAsnLe 367
1540 GGCCGCTCACTCTTACAGTTCACACCTCGTCCGCTGATGCCAACCCT 1589
367 uGluAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaPheProA 384
1590 GGAAGACTTTCCTTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1639
384 laglylleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400
1640 CAGGATCACTGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1689
401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysProLe 417
1690 ACGGATGCCAGCTCTGTGGACACCCAGCTGGCAACTGCAAGCCCTC 1739
417 uGlnLysAspLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGln 434
1740 TCAGAGGACAGCTCCCTCAGAACCTCTCTCTCTCTCTCTCTCTCTCTCT 1789
434 erLeuThrAlaLeuThrLeuGlnValGluAsnLysSerAsnThrValGln 450
|||||

1790 GCCTCAGCGCGCTGACCTTACAGGTGGAGAACATCTCCAAACCGCTCCAG 1839
451. GlnLeuLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLysAs 467
1840 CAGCTCTGCTCTCCAAAGGTCTGTGCGCAGAAAGAGTGTCAAGAA 1889
467 nLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerProG 484
1890 CPTCGGTGTCAGGACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1939
484 lucGlySerGlyTyrSerAlaGlnProAlaGlyThrProLeuSerGluPro 500
1940 AAGGAGGCGGTACTCAGCCGAGCCCGGCGCACACCGCTGTCAAGCCG 1989
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGluAlaAspTy 517
1990 CCGAGCAGCAGCCACAGTCCAGCATCGGAGCGCGCAGGAGCGCGACTA 2039
517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCysAsnG 534
2040 CTTGAGCGGCTCCGAGGACCCACTGGAGCGAGCTTCTTACTGCAACC 2089
534 lnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550
2090 AGCCCGGTGGCAGCCCTGCGCAGGTCAACAGCAACTCGAAGCCCAAGCC 2139
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567
2140 GATCTGCTGCTCCACTGTTCTGTGACCTCTCTCTGAGCAGCATGTCCACAA 2189
567 sSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584
2190 ATCTGACGACTCTTCCAGAGCCTACAGCGAGTCTGCCGCTCGACAGCT 2239
584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuLeuSer 600
2240 TCTCAAAGTCTGTGGCGGTGAGCGGAGTGTCCGCGGCTGCTGCTCAGC 2289
601 AlaLeuAlaGlnGluAspLeuAlaSerGluIleLeuGlyLeuGlnGluAl 617
2290 GCCTTGGCAGGAGGAGCTTGGCTCCGAGATCTTGGGCTGCGAGGAGC 2339
617 alledGlyGluLysAlaAspLysAlaTyrAlaGluAlaProSerLeuVal 634
2340 CATCGGTGAGAAGGCCGACAAAGCTTGGGCTGAAGCAGCAGCAGCTGTGCA 2389
634 ysAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLeu 650
2390 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2439
654 AspSerValAlaLysSerAlaTyrProArgProGlyGluProGluAlaLe 667
2440 GACTCTGTGGCAGAGTGTGCTGGCCCGGCTGGGAGCCCGAGGCCCT 2489
667 uProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684
2490 GCCCGACTCTTGGCAGGAGGCGGCAATGCAAGGACTTCAGCC 2539
684 rodGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700
2540 CAGGGCTGTTGAAGACCTTCTGCTGGCTTCTGCTACGCTACGCCCAA 2589
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717
2590 AAGACAACCTGCTCTCTCTCTTGTGTAACCAAGCCACCCCTGGGGTTC 2639
717 calaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734
2640 TGCTCCAGACCCCTACAGCAGCTTTGACTGTTTCCGGACACACCG 2689
734 laaLeuSerAlaAspSerAlaAsnProPheAlaTyrProGluGluAsn 750
2690 CTGCCAGCTCAGCGGACAGCGCCACACCTTTCCTTGGCTGGCGAGGAAAC 2739

1051	GLYLEUPROALMECYSTRINASERSEINTHIALAUSERGILUPROK	1067
3640	GGGCTCCCGAGATATGTACTGCTTCTCCAGGCCCTGATAGCCCG	3639
1067	gthProGlyProProGlyLeuThrThrProAlaProProAspLysL	1084
3690	CACGCCCGAGACCCCGAGGCTCGACACACCCTCGACCCCAAAAC	3739
1084	euGIYGLYLYSGINHRALALAPHELYSERTLYYHARVALGILYS	1100
3740	TGGGGGACAGACGAGGACCGCCCTTCAAGTGGGACAGGGGTGGAAAG	3789
1101	ProSerProLYsAlaAlaSerSerProSerAspProAlaAlaLeuProVa	1117
3790	CGCTCAACCCAAAGGTGGCTCCACACCACGAAACCGGGCGCCGTCTG	3839
1117	AlaSerAspSerSerProMetLysSerLYsLYsGLuThrAspSerP	1134
3840	GGCTCCCGACACACCCCGATGGCGCTCCAAACCAAGGAGACACACAC	3889
1134	roSerThrProGlyLYsAspGlnArgSerMetLLeuArgSerArgThr	1150
3890	CCAGCAAGCTGGCAAGACACACAGCTCTCATGATCTTGGTCAAGCAC	3939
1151	LYsThrGlnGlnLLeuPheLysSerLYsAspArgArgProSerGlnLYsR	1167
3940	AAACCCAGAGATCTTCCACTCCAAAGGGGAGAGCCCTGTGAGGGCCG	3989
1167	glnProAspCYsArgArgAlaThrLYsLYsLeuLeuAspAsnSerLsleuP	1184
3990	GGTCCCAATGCTGCTGGCACCAAGAGACTCTTCGAAACAGCCACTGG	4039
1184	roAlaThrPheLYsValSerSerSerProGlnLYsGlnLYsArgValSer	1200
4040	CGCGCAATTCAAAGGTCTCCAGACGCCCAAGAGAGGCGCAGGGTACG	4089
1201	GlnArgAlaArgValProLYsProGlyAlaLeuLYsLYsLeuSerAspAr	1217
4090	CAGGGGCAAGGTCTCCCAACCTGGTCCAGACGACGAACTCTTGACCG	4139
1217	gProLeuHAlaLeuLYsArgLYsSerAlaPheMetAlaProValProT	1234
4140	GGCCCTCCAGGCGCAAAAGAGAGTGGCGCTTCAAGGGCGGGTCCCA	4189
1234	hLYsLYsArgAsnLeuValLeuArgHLSLYsSerSerSerSerAsn	1250
4190	CCAGAGAGCGGAACTGGTCTTCCGGCACGGCACGACGACGACGAC	4239
1251	AlaSerLLeuMetGILGluMetGILYArgArgArgGILLeuArgVALProPr	1267
4240	GGCACTGCATGGGGAGATGGGAGAGAGAGAGGCTTAGGGTTCCCC	4289
1267	oProSerSerArgLYsLeuLeuProArgLYsProSerProProArgA	1284
4290	ACCCCTTCAGAGAGATGTCCTTCCCAAGAAAGCAACCCACACAGGG	4339
1284	lAlaMetAlaSerLeuProLysSerSerHisProArgArgProPheLeu	1300
4340	CATGGGAGACCTGGCAAAAGGTCCACCCCGGAGAACCCCAATTCGTG	4389
1301	ProGlnAlaArgLeuSerSerAlaAlaPheGlnGILValMetLYsThrLYsVA	1317
4390	CGTCAACTGGCTCGCGAGCGCTTCCAGGGGGCCATGAAGACAAAGG	4439
1317	lLeuProProArgLYsGLYArgGILYLeuLYsLeuGILuAlaLLeuAlaGlnL	1334
4440	GGTGGCAACCCCGAAGGGCGGGGCGCTCAACCTGGAAGACCACTGTGAA	4489
1334	YsLLeuThrSerProSerLeuLYsLYsPheAlaCYsLYsLAlaProGlyALA	1350
4490	AAGATCACTGGCGACGCTCAAGAGTTCCGATGTAAAGCGCAGGGGCG	4539
1351	serProGlyAsnProLeuSerProSerLeuSerAspLYsAspArgGILYLe	1367

```
|||||
4540 TCTCTGGTAATCTCTGAGCCCATCCCTTTCCGACAAAGACCGTGGGCT 4589
|||||
1367 uLysgLyAlaGlySerProValGlyValGluGlyLeuValAsnV 1384
|||||
4590 CAGGGTGCTGGGGCGACCCACATGGGGTGGAGAGGGCTGGTAATG 4639
|||||
1384 alGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCysArg 1400
|||||
4640 TGGGACCGGGGAGAGCTCCCAACTCTGGGGCTGATCCGTTATGCAGA 4689
|||||
1401 AsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLysLe 1417
|||||
4690 AATCCACCAACAGATCTTAAAGGCAACCTCATGACAGTAAAGAACT 4739
|||||
1417 uSerSerThrAspCysPheLysThrGluAlaPheThrSerProGluAlaL 1434
|||||
4740 GTCTCTTACTGACTGTTTCAAAACCGAGGCTTCATCCCGGAGGCC 4789
|||||
1434 euGlnProGlyThrAlaLeuAlaProLysLysArgSerArgLysGly 1450
|||||
4790 TCCAGCCTGGGGGAGCTGCCCTCGGCGCTAAGAGAGGAGCCGGAAGGC 4839
|||||
1451 ArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGlnLysArgProTy 1467
|||||
4840 CGGGCAGGGGCCCATGGACTCTCCAAAGGCCCGCTGGAGAGCGGCCCTA 4889
|||||
1467 rLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyThrG 1484
|||||
4890 TCTTGGCCGGCTCTCTCTGATCCCGGAGACAGGGCCAGTGGCACAC 4939
|||||
1484 lngLyAlaSerGluAspAsnSerGlyGlyGlyLysLysProLysMet 1500
|||||
4940 AAGGGCCAGTGGAGCACTCTGTGGAGGAGGCAAGAGCAAGATG 4989
|||||
1501 GluGlnLeuGlyProAlaSerGlnProProGluGlyArgProCysGlnPr 1517
|||||
4990 GAGAGCTGGGCCCTGCTCCAGCCCGCGAGGGGCGAGGCCCTGCCAGCC 5039
|||||
1517 oGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTyrs 1534
|||||
5040 CCAGCAAGGCGCAGAAACAGCAGCCAGCCACCACTACAGCACTATT 5089
|||||
1534 erLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSerSer 1550
|||||
5090 CCAAGCGAAGCGCTCCTCGGGGCGGGCCGAGAACACCACTCTCTCA 5139
|||||
1551 ProCysLysGlyAlaGlnLysArgArgGlnGlnGlnValLeuProLe 1567
|||||
5140 CCTGTAGGGGCGTCCAGCGCAGCAGCAGCAGCAGCAGTGTGCCCT 5189
|||||
1567 uAspProAlaGluProGluLeuArgLeuLysTyrrLysSerCysLysA 1584
|||||
5190 GGATCCCGCAGAGCTGAAATCCGCTCAAGTACATTTCTCTTCCACAG 5239
|||||
1594 rGluArgSerAspSerArgThrProAlaPheSerProPheValArgVal 1600
|||||
5240 GGTGTAGGTGACAGAGCGAGCCCGCTCTCTCCCTTCGTCGGGGTG 5289
|||||
1601 GluLysArgAspAlaPheThrThrLysCysThrValValAsnSerProG 1617
|||||
5290 GAGAGCGAGAGCGCTTCCACCATATGCATGTGTCAACTCCCTCGG 5339
|||||
1617 yAspAlaProLysProHisArgLysProSerSerSerAlaSerSerSerS 1634
|||||
5340 AGATGCGCCCAAGGCCCAAGAGAGGCTCTCTCTCTCTGCTCTCTCT 5389
|||||
1634 erSerSerSerSerPheSerLeuAspAlaGlyAlaSerLeuAlaThr 1650
|||||
5390 CATCTGCTGCTCTCTCTCTCTGATGAGCGGGGCTCCCTGGCCACA 5439
|||||
1651 LeuProGlyLysSerIleLeuGlnProArgProSerLeuProLeuSerSe 1667
|||||
```

```
5440 CTCCCTGAGGCTCCATCTCTGACGCCGCGCCCTCTCTGCCCCCTCTCC 5489
1667 rThrMerHisLeuGlyProValValSerLysAlaLeuSerThrSerCysL 1684
5490 CAGGATGACTTGGGCGCTTGTTTCCAGGCGCTGAGTACCTCTTGCC 5539
1684 euValCysCysLeuCysGlnAsnProAlaAsnPheLysAspLeuGlyAsp 1700
5540 TTGTTTGTGCTGCTCTGCCAAACCGGCGCAACTTCAAGGACCTTGGGG 5589
1701 LeuCysGlyProTyrrProGluHisCysLeuProLysLysLysProLy 1717
5590 CTCTGTGGGCGCTACTACCTGACCTGCTCCCTCCCAAAAGAACCCCAA 5639
1717 sLeuLysGlnLysValArgProGluGlyThrCysGluGlnAlaSerLeuP 1734
5640 ACTCAGAGAGAGGTGGGCCAGAGGACCTGTGAGGAGGCGCTCGCTGC 5689
1734 roLeuGluArgThrLeuLysGlyProGluCysAlaAlaAlaThrAla 1750
5690 CGTTGAGAGAACTCAAGGTCCGAGTGTGAGTGTGAGTGTGAGTGTGAG 5739
1751 GlyLysProProArg 1765
5740 GGAAGCCCCCCCCAGG 5754
seq_name: gb_htg:HS253P07
seq_documentation_block:
LOCUS HS253P07 160544 bp DNA HTG 20-APR-2001
DEFINITION Homo sapiens chromosome 17 clone PAC RPI-1 253P07 map 17p11.2, ***
SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION AL354000.6 GI:13752103
VERSION AL354000.6 GI:13752103
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 160544)
AUTHORS Ramser, J., Langer, I., Steffens, C., Klein, M., Borzym, K., Junker, E.,
Marquardt, I., Moll, K., Radelof, U., Francis, F., Seranski, P.,
Poustka, A., Reinhardt, R. and Lehrach, H.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 160544)
AUTHORS MOLGENR.
TITLE Direct Submission
JOURNAL Submitted (28-APR-2000) MOLGENR, Abt. Lehrach, Max Planck Institut
Fuer Molekulare Genetik, Innestrasse 73, Berlin, 14195 Germany
On May 14, 2001 this sequence version replaced gi:9931105
gi:10280527.
contig 01 1..86093
contig 02 86194..160544
Clone received from the Resource Centre of the Human Genome Project
at the Max-Planck-Institute for Molecular Genetics.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 86093: contig of 86093 bp in length
* 86094 86193: gap of 100 bp
* 86194 160544: contig of 74351 bp in length.
Location/Qualifiers
1..160544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
FEATURES
source
```



```
/clone="PAC RPCI-1 281113"  
/clone_lib="RPCI1,3-5 Human PAC library, originating  
institute: Roswell Park Cancer Institute, creator: Pieter  
de Jong, P. Ioannou"  
/note="region between markers D17S842-D17S953"  
misc_feature  
1  
/note="77_end:PAC RPCI-1 253P07"  
misc_feature  
160544  
/note="SP6_end:PAC RPCI-1 281113"  
BASE COUNT 35434 a 44179 c 44032 g 36798 t 101 others  
ORIGIN  
  
alignment_scores:  
Quality: 9191.00 Length: 1758  
Ratio: 5.249 Gaps: 3  
Percent Similarity: 99.602 Percent Identity: 99.545  
  
alignment_block:  
US-09-508-821B-6 x HS253P07/rev  
  
Align seg 1/1 to reverse of: HS253P07 from: 1 to: 160544  
  
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTyr 17  
109850 ATGCAAGCTTTCGAGAAAGGTGGTTTCATGGCAACACACAGACTA 109801  
17 GlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyrHisArgLys 34  
109800 CCAGCAGACCTCGCAGAAACATCACCCCTAGAGAAATACAGGAGCCGA 109751  
34 ArgGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTyr 50  
109750 GTCAGAGCCGGGGCTAAGCTGACGACCGGAGCGGCTGCTCCGCAAGACTAT 109701  
51 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyTyrHisProse 67  
109700 TATAACCCGAGGCTTACCCGAGCTATGAGGGGTGGGCGGCGACGCCCTTC 109651  
67 GlnTyrAlaAlaAlaAlaAlaAlaAspLysTyrHisArgLysTyr 84  
109650 TGGCAGCTGACCGCGGTGGCCCGGACAGTACACCGAGGAGCAGCAAG 109601  
84 LeuProThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100  
109600 CCGTCCCAACAGCAAGGCTGACGAGGAGCGCGGCTTCCCTGCGCTAC 109551  
101 GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlnGlnGln 117  
109550 GCGGTCCAGGACAGCAAGCCCTTACCCAGGCGGCTATGCTGGTGAAGAG 109501  
117 GlnGlnAlaThrPheGlyAlaProGlnProProProProGlnProGlnPro 134  
109500 CCTTCAAGGCTTGGGGGGCCCGCCACAGCACCACCCCGCAGCGGAGCCAC 109451  
134 LeuProAlaGlyValAlaLysTyrAspLysLysLysLysLysLysLys 150  
109450 TACCTGCAAGGGGTGGCCAGTATGATGAACTTGATGAAAAGACAGCA 109401  
151 ValProProSerArgGlnTyrAlaGlnGlnGlnGlnGlnGlnGlnGln 167  
109400 GTGCCCCCAGCAGGAGCAGTATGACAGGAGGAGGAGGAGGAGGAGGAG 109351  
167 GlnHisSerLeuHisValGlnGlnProProProProProGlnGlnProLeu 184  
109350 GACTCACTCCCTGCACTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 109301  
184 LeuTyrProLysLeuGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 200  
109300 CATACCCCAAGGCTCCAAAGCAGCAAGGCTGAGACGACATCTCCCTCC 109251  
201 LeuProPheProGlnGlnGlnTyrHisPheProGlnHisSerGlnSerPhe 217
```

108350 CCGAGCAGCAGCCACAGTCCAGCATCGGAGGCCAGGAGGCCGACTA 108301
517 rLeuSerGlySerGluuaspProLeuGluArgSerPheLeuThrCysAsnG 534
108300 CCGAGCAGCAGCCACAGTCCAGCATCGGAGGCCAGGAGGCCGACTA 108251
534 lnaLaArgGlySerProAlaAlaGValAsnSerAsnSerLysAlaLysPro 550
108250 AGGCCGCTGGCAGCCCTGCGAGGCTCACAGCAACTCGAGGCCAAGCCC 108201
551 GluSerValSerThrCysSerValThrSerProAspMetSerThrLy 567
108200 GAGTCGCTGCCACCTCTCTGACTCTCTGACGATCTCTGACGATCTCCACAA 108151
567 sSerAspSerPheGluSerLeuHisGlySerLeuProLeuAspSerP 584
108150 ATCTGACGACTCTCTCAGAGCTACAGGAGCTGCGGCTGACAGCT 108101
584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuSer 600
108100 TCTCAAAGTCTGCGGGGTGAGCGGACTGTCCGGGCTGCTCAGC 108051
601 AlaLeuAlaGlnGluaspLeuAlaSerGluLeuGlyLeuGlnGluAl 617
108050 GCCCTGGCAGAGGAGCTTGGCTCCAGATCTCTGGGCTGCGAGAGC 108001
617 alleGlyGluLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 634
108000 CATCGGTGAGAGGCCGACAAAGCTTGGGCTGAAGCACCCAGCTGGTCA 107951
634 ysAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLeu 650
107950 AGGACAGCAGCAAGCCACCTCTGCTGGAGAACACAGCGCTGCTG 107901
651 AspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAlaLe 667
107900 GACTCTGTGCCAGAGCTGGTGGCCCGGCTGGGAGCGGAGGCCCT 107851
667 uProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684
107850 GCCCGACTCTTGCAGCTGACAAAGGGGGCAATGCCAAGGACTTCAGCC 107801
684 roGlyLeuPheGluaspProSerValAlaPheAlaThrProAspProLys 700
107800 CAGGCTGTGTTGAGACCCCTTCGGTGGCTTCGCTACGCTGACCCCAA 107751
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717
107750 AAGACAACGTGTCTCTCTCTTGGTACCAAGCCACCCCTTGGGGTTCC 107701
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThrA 734
107700 TGCTCCAGACCCCACTACAGAGCTTTTGAATGTTTCCCGGACACACCG 107651
734 laAlaSerSerAlaAspSerAlaAsnProPheAlaTrpProGluGluAsn 750
107650 CTGCCAGCTCAGCGGACAGCGCCACCCCTTGGCTGGCCAGAGGAAAC 107601
751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrLy 767
107600 CTGGGGGATGCTTGTCCAGGTGGGATTCACCCCTGGGAGCTTAOCAA 107551
767 sGlyLeuGluGlnGlyGlyLysAlaSerAspGlyLysSerLysGlyAspT 784
107550 GGGCTGGAGCAGGGTGGAGGCCCTCAGATGGCATCACAAAGGGGACA 107501
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGluLysAspProProGly 800
107500 CCCATGAGCTTGGGCTGCTGGCTTCCAGAGGAGGAGGCCCCCTGGG 107451
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGlyG 817
107450 GAGAGGTGGCTGCTTGGCCCGGGGACTTCAAGCGAGGAGGTGGGTGG 107401

817 yValLysGlnGluAlaGlyGlyLeuLeuGlnCysProGluValAlaLysA 834
107400 GGTGAAGAGAGAGCAGTGGCTGCTGAGTGCCTCCCGAGGTGGCCAAAG 107351
834 laAspArgTrpLeuGluaspSerArgHisCysCysSerThrAlaAspPhe 850
107350 CTGACCGGTGGCTGGAGGACAGCGGCTGTGTGTTCACCGCGGACTTC 107301
850 GlyAspLeuProLeuLeuProThrSerArgLysGluaspLeuGluAl 867
107300 GGGGACCTTCCACTGCTGCCACCACAGCAGAAAGAGGAGGAGCTGGAAGC 107251
867 aGluGluLeuTySerSerLeuCysGluLeuLeuGlySerProGluGlnA 884
107250 TGAGAGGAGTACTCTCTCTATGTGAGCTTCTGGGCGAGCCCGGAGCAGA 107201
884 rProGlyMetGlnAspProLeuSerProLysAlaProLeuLeuCysThr 900
107200 GGCCTGGCATGAGGACCGGCTGTACCCAAAGGCCCTCATCTGTCACC 107151
904 LysGluGluValGluGluValLeuaspSerLysAlaGlyTrpGlySerPr 917
107150 AGGAGGAGGTGGAGAGGTGTGACTTCCAGGCCGCTGGGCTGCTCC 107101
917 oCysHisLeuSerGlyGluSerValLleLeuLeuGlyProThrValGlyT 934
107100 GTGCCACCTCTCAGGGGAGTGTGCTCATCTGTGGGCTTACAGTGGCA 107051
934 hrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950
107050 CGGAGTCAAAGGTCAGAGCTGGTGTGAGTCTCTCTCACACATGAAG 107001
951 ProGlyGluGluGlyProAspGlyGluArgAlaProGlyAspSerThrTh 967
107000 CCAGGTGAGAGGGGCTCTGATGGGAGCGAGCTCCAGGGGATTCACAC 106951
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984
106950 CTGGACGCTCTCTGCGCCAGAAAGAGCTGTGCCAGGGGCAAAAGCTTACG 106901
984 laProileAlaLysLysGluProValProArgGlyLysSerLeuArgSer 1000
106900 CGCCCATCCCAAGAAAGAGCTGTGCCAGGGGCAAAAGCTTACGAGC 106851
1004 ArgArgValHisArgGlyLeuProGluAlaGluaspSerProCysArgAl 1017
106850 CGTCCGGTGGCAGCGGGGCTGCCAGCGGAGGACTCCCATGAGGAGC 106801
1017 aproValLeuProLysAspLeuLeuLeuProGluSerCysThrGlyProP 1034
106800 ACCAGTGTGCCAAAGAGCTTGTCTCCCTGAATCCTGCACAGGGCCCTC 106751
1034 roGlnGlyGlnMetGluGlyValAlaGlyAlaProGlyArgGlyAlaSerGlu 1050
106750 OCCAGGACAGATGAAGGGCTGGAGCCCGCAGGCGGGGGCTTCGGA 106701
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067
106700 GGCTCCCGCAGGATGTACTGTCTCTCACGGCTTGTGAGGAGCCCG 106651
1067 gThrProGlyProGlyLeuThrThrProAlaProAspLysL 1084
106650 CAGCGCCGGACCCCGAGGCTGACCCACCCCTGCACCCCGCAGACAAAC 106601
1084 euGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100
106600 TGGGGGCAAGCAGCAGCGGCTTCAAGTCGGGCAAGCGGTGGGGAAG 106551
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117
106550 CCTCACCCAGGCTGCTTCCAGCCCGCAGCAACCCCGCCCTGCTCTGT 106501

1117 AlaSerAspSerSerProMetGlySerLysThrLysGluThrAspSer 1134
106500 GGCCTCCGACAGCAGCCGATGGGCTCCAGAACAGAGAGAGAGCTCAC 106451
1134 roSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerThr 1150
106450 CCAACAGCCCTGGCAAGACCGCTCCAGATGATCTTGCGACAGCCACC 106401
1151 LysThrGlnIlePheMetSerLysArgArgArgProSerGluGlyArg 1167
106400 AAAACCCAGAGATCTTCACCTCAAGGCGAGGCGCTCGAGAGCGCG 106351
1167 GluProAsnGlyArgAlaThrLysLysLeuLeuAspAsnSerHisLeuP 1184
106350 GCTCCCAACCTCCGCTCCACAGAGAGCTCTCGACACAGCCACTGCG 106301
1184 roAlaThrPheLysValSerSerSerProGlnLysGluGlyArgValSer 1200
106300 CCGCCACATTCAGAGTCTCCAGAGCCCGAGAGAGAGAGAGGAGTGGC 106251
1201 GlnArgAlaArgValProLysProGlyAlaGlySerLysLeuSerAspArg 1217
106250 CAGGGGCAAGGGTCCCAAACTGCTGAGGAGGAGGAGGAGGAGGAGG 106201
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValProT 1234
106200 GCCCTCCATGCGCTCAAAAGAGAGTGGCTTCATGAGGCGCGCTCCCA 106151
1234 hrLysLysArgAsnLeuValLeuAlaGlyHisGlySerSerSerSerAsn 1250
106150 CCAAGAGGCGAGACCTGCTTCGAGAGCCGACAGCAGCAGCAGCAGCAG 106101
1251 AlaSerAlaMetGlyGlnMetGlyArgArgArgGlyLeuArgValProp 1267
106100 GCCAGTGGCAATGGGAGAGATGGAGAGAGAGAGAGAGAGAGAGAGTCCC 106051
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProProArg 1283
106050 CACCTCTTCAGAGAGATGCTCTCCAGAGAGAGAGAGAGAGAGAGAG 106001
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProProPheL 1300
106000 GCATGGCGAGCGCTGCGACAAAGCTCCACCCCGGAGAGAGAGAGAGAG 105951
1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLys 1316
105950 TGCTCAAGCTCGCTCGGAGAGCTTCAGAGGAGGAGGAGGAGGAGGAG 105901
1316 sValLeuProProArgLysGlyValArgGlyLysLeuGluAlaIleValG 1333
105900 GGTCGTCCAGCCCGAGAGAGCGGAGGCTGAGAGCTGAGAGCAATCGTGC 105851
1333 LysLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349
105850 AGAAGATCACTCGCGCCAGCTCAGAGAGTGGCATGTAAACGCGAGAGG 105801
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgGly 1366
105800 GCCTCTCTGATATCTCTGAGCCCATCTCTTCGACAGAGAGAGAGAGAG 105751
1366 LysLeuLysGlyAlaGlySerProValGlyAlaGluGluGlyLeuValA 1383
105750 GCTCAAGGGTGTGGGGGAGCCAGAGTGGGGGTGAAAGAGAGCTGGTAA 105701
1383 snValGlyThrGlyLysLysLeuProThrSerGlyAlaAspProLeuGly 1399
105700 ATGTGGGACCGGGGAGAGCTCCCACTCTGGGGGTATCCGTTAAGC 105651
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLys 1416
105650 AGAATCCACCAACAGATCTTAAAGGCAACTGATGAGAGAGTAAAGAA 105601
1416 sLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGluA 1433
105601 ACTGCTCTCTACTGACTGTTTCAAAACGAGAGGCTTCACATCCCGGAGG 105551
1433 LysLeuGlnProGlyLysThrAlaLeuAlaProLysLysArgSerArgLys 1449
105550 CCTGCAAGCTGGGGAGAGCTGCCCTGGGCGCTTCAAGAGAGAGAGAGAA 105501
1450 GlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgPr 1466
105500 GCGCGGAGAGGGGCCATGAGACTCCAAAGAGCCGCTGAGAGAGAGGCC 105451
1466 cTyLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyT 1483
105450 CTATCTGAGCCCGGCTGCTGCTCTACTCCCGAGAGAGAGGCTGAGTGA 105401
1483 hrGlnGlyAlaSerGluAspAsnSerGlyGlyGlyLysLysProLys 1499
105400 CACAAGGGCCAGTGAAGACAACTGTGTGAGAGAGGCAAGAACCAAG 105351
1500 MetGluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysGly 1516
105350 ATGAGAGAGCTGGGCTGGCTCCAGCCCGGAGGAGAGAGAGAGAGAG 105301
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnThrSerSerT 1533
105300 GCCCGAGAGAGGCGCAGAGAACAGCAGGAGCAGACCAACTAGCAGAGCT 105251
1533 YrSerLysArgLysAlaGluThrArgGlyValAlaLysAsnThrThrSer 1549
105250 ATTCCAGAGGAGAGCGCTCCTACTCGGGCGGAGGAGAGAGAGAGAG 105201
1550 SerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuP 1566
105200 TCACCTGTGAGGGGCGCTGCAAGAGAGAGAGAGAGAGAGAGAGAGCT 105151
1566 cLeuAspProAlaGluProGlnIleArgLeuLysYrIleSerSerCysL 1583
105150 CCGTGTATCCCGAGAGCTGAGAAATCCGCTCAAGTCAATTCCTGTGCA 105101
1583 YrArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599
105100 AGCGCTGAGGTGAGAGAGCGGAGAGCCCGGCTTCACACCTGCTGGG 105051
1600 ValGluYrArgAspAlaPheThrThrIleCysThrValAlaAsnSerPr 1616
105050 GTGAGAGAGAGAGAGCGCTTCACACATATGAGTGTGCAACTCCCC 105001
1616 cGlyAspAlaProLysProHisArgLysProSerSerSerAlaSerSerS 1633
105000 TGGAGATGCGCCCAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 104951
1633 eSerSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAla 1649
104950 CCGTCACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 104901
1650 ThrLeuProGlyLysSerLysLeuGlnProArgProSerLeuProLeuSe 1666
104900 ACHCTCCGAGAGCTCATCTGCAAGCGGAGGCTCTGCTGCTGCTGCT 104851
1666 rSerThrMetHisLeuGlyProValValSerLysAlaLeuSerThrSerC 1683
104850 CTCACAG 104801
1683 YrLeuValCysCysLeuCysGlnAsnProAlaAsnPheLysAspLeuGly 1699
104800 GCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 104751
1700 AspLeuCysGlyProYrThrProGlnHisCysLeuProLysLysPr 1716
104750 GACCTGTGTGGGCGCTTACCTTACCTTACCTTACCTTACCTTACCTT 104701
1716 cYrLeuLysGluLysValArgProGluGlyThrCysGluAlaSerL 1733

9640 CCAAGACACCTGCGAGGAACATCAAGCCTAGAGAAATTACAGGACGCGA 9669
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTyr 50
9690 GTCAAGCGCGGCTTAAGCTGAGACCGGACCGGCTGCTGCGCAAGGAGCTAT 9739
51 TyrAsnProGlnProTyrProSerTyrGlnGlyValAlaGlyThrProse 67
9740 TATAACCGGAGCTTACCGGAGCTATGAGGCTGGCCGTGGCAGCGCTTC 9789
67 rGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84
9790 TGGCACTGCAAGCGCGGAGGCGCGGACAGATACACCGAGGACAGAG 9839
84 LaleuProthrGlnGlnLysLeuGlnGlyArgProAlaIleProGlyTyr 100
9840 CCTGCCACACAGCAGCGCTGAGAGGAGCGCGCTTCCTGCTGCTG 9889
101 GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlnGlnLys 117
9890 GAGCTCCAGGACAGAGCGCTTACCGAGCGCGCTATGCTGGTGAAGAG 9939
117 rLeuGlnAlaTyrGlyAlaProGlnProProProProGlnProGlnProL 134
9940 CCTTCAGGCTTGGGGGGGCCCGACAGCACACCGCCACAGCGGACGAC 9989
134 euProAlaGlyValAlaLysTyrAspGlnAsnLeuMetLysGlyThrAla 150
9990 TACCTGCAAGGCTGGCCAGATGATGAGAACTGATGATAAAAGACAGCA 10039
151 ValProProSerArgLysTyrAlaGlnGlnGlnValAlaGlnValProPheAr 167
10040 GTGCCCCCAGGACGAGATGACAGAGGCGCGCCAGGCGCTTTCG 10089
167 gThrHisSerLeuHisValGlnGlnProProProProGlnGlnProLeuA 184
10090 GACTCACTCCCTGAGCTCCAGACAGCCACCGCGCCCGACAGCGCTTCG 10139
184 LalyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleHisSerPro 200
10140 CAAACCCAGGCTCCAAAGGAGAGCTGCGAAGACATTCCTCCCTCCCT 10189
201 LeuProPheProGlnGlnLysThrHisPheProGlnHisSerGlnSerPhePr 217
10190 CTGCCCCCTTCCCGAGGCTACCCATTCCTCAGCATTCACCATCTCTCC 10239
217 oThrSerSerThrTyrSerSerSerValGlnGlnGlyGlyGlnGlnAla 234
10240 CACTCTCTCACTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10288
234 LserTyrLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250
10289 ACTCCTATTAAGATTGCACAGCACGAGCTGCCAGGCCCATGACAGGCG 10338
251 LeuThrAlaSerSerSerLeuAlaProGlnGlnArgValGlnAsnLeuH 267
10339 CTGACTGCGAGCTCCAGCTCGGCGCGGCGAGCGGCTCCAGATCTTCA 10388
267 salatyrglnserGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGln 284
10389 TGGCTACAGTGGGGCGGCTCAGTATGACACAGACAGACAGACAGCGC 10438
284 LnglnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
10439 AGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 10488
301 GluThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlnGln 317
10489 GAHACCTCATTTACCAAAACCTCGCCAGATGACAGCAGCAGCAGCAGCA 10538
317 nGlyGlnGlyTyrCysGlnProAspAlaValAlaArgThrProGlnGln 334
10539 AGCCAGGAGCTACTGCGAGCGGAGCGACCGCTCGGAGCCCGCAGAGCACT 10588
334 yTyrGlnGlnProPheSerProSerSerSerHisSerProAlaArgSerVal 350
10589 ACTACAGACCTTCAACCCCGACGCTCCAGCAGCAGCAGCAGCAGCAGCAG 10638
351 GlyArgSerProSerTyrSerSerThrProSerProLeuMetProAsnLe 367
10639 GAGCGCTACCTTCTACAGTTCCACACCTGCGCGCGCTGAGGCAAACT 10688
367 uGlnAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaIleProA 384
10689 GAGAACTTCCCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 10738
384 lAglyTleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400
10739 CAGGATCATGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 10788
401 ThrAspAlaThrSerSerValAspThrAlaAlaGlyAsnAspLysProLe 417
10789 ACAGATGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 10838
417 uGlnLysAspLysLeuProGlnAsnLeuSerAspLeuSerLeuGlnS 434
10839 TCAGAAAGACAAAGCTCCCTGAGAACCTGCTGAGATCTCAAGCTGACAG 10887
434 eLeuThrAlaLeuThrLeuGlnValGlnAsnLleSerAsnThrValGln 450
10888 C.CTCAAGCGCGCTGACCTCAAGGTTGGAGAACATCTCAACACCTCCAG 10937
451 GlnLeuLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLysAs 467
10938 CAGCTCTCTCTTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10987
467 nLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerProG 484
10988 CCGTGCTCAGAGACCCAGCAGCAGCATAAAGCAGCAGCAGCAGCAGCAG 11037
484 lGlySerGlyTyrSerAlaGlnProAlaGlyThrProLeuSerGlnPro 500
11038 AAGGAGCGGCTACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 11087
501 ProSerSerThrProGlnSerThrHisAlaGlnProGlnGlnAlaAspTyr 517
11088 CCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 11137
517 rLeuSerGlySerGlnAspProLeuGlnArgSerPheLeuTyrCysAsnG 534
11138 CCGTACAGGCTCCGAGGAGCCACAGGAGCAGCTTCTCTACTGCAAC 11187
534 lAlaIarGlySerProAlaArgValAsnSerHisSerLysAlaLysPro 550
11188 AGGCCCTGCGACCCCTGCGAGGCTCAACAGCACTCGAAGCGCAAGCCC 11237
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567
11238 GAGTCCGCTGACCTGTTCTGTGACTCTCTCTGAGCAGACAGTCCACCA 11287
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584
11288 ACTGACAGCTCTTCCAGAGCTTACAGGCACTGCGCGCTCCAGAGCT 11337
584 hSerLysPheValAlaGlyGlnArgAspLysArgLeuLeuLeuSer 600
11338 TCTCCAGTTCTGCGGCGGTGAGGAGGAGCTGCGCGGCTCTCTGAG 11387
601 AlaLeuAlaGlnGlnAspLeuAlaSerGlnLleLeuGlyLeuGlnGln 617
11388 GCGCTGCGACAGGAGAGCTGAGCTCCGAAATCTTGGAGCTGCGAGAGAG 11437
617 alIleGlyLysLysAlaAspLysAlaTyrAlaGlnAlaIleProSerLeuVal 634
11438 CACGCTGAGAGGCGCAGCAAAAGCTTGGGCTGAGAGACACCGACTGATCA 11487

13288 CCAAGAGCGGACCTGGCTTGGCGAGCCGACAGACGACGACGAC 13337
 1251 Alaser .Alametgylumetgylarargargylleuargvalprop 1267
 13338 GCCAGTGGCAATGGGGGAGATGGGAAGGAGAGAGGCTGAGAGGTGCC 13387
 1267 roProSerSerArggLyCysLeuLeuProArgLysProSerProArg 1283
 13388 CACCTCTTCAAGAGAGATGTCTTCCCAAGAAAGCAAGCCCAAGG 13437
 1284 AlametAlaserLeuProGlnSerSerHisProArgArgProPrope 1300
 13438 GCAATGGCGAGCTGGCCCAAAAGCTCCACCCCGAGAGCCCGAGTGC 13487
 1300 eufProGlnAlaArgLeuSer .AlaAlaPheGlnGlyAlaMetLysThr 1316
 13488 TGCCTCAAGCTGCGCTCTCGGGAGCTTCCAGGGGGCCATGAGACCA 13537
 1316 sValLeuProProArgLysGlyValGlyLeuLysLeuGlnAlaLeuVal 1333
 13538 GGTCTGCGCCACCGGAGAGGCGGGGCTGAGAGCTGAGAGCCATGTC 13587
 1333 InLysLleuThrSerProSerLeuLysLysPheAlaCysLysAlaPro 1349
 13588 AGAAGATCACTCGCCGACCTCAAGAGTTGCGATGTAAGAGCGCAAG 13637
 1350 AlaserProGlyAsnProLeuSerProSerLeuSerAspLysAspArg 1366
 13638 GCTCTCTCGTGAATCCTCTGAGACCATCTTCCGAAAGAGCCGTGG 13687
 1366 YLeuLysGlyAlaGlyLysSerProValGlyValGlnGlnGlyLeuVal 1383
 13688 GCTCAAGGGTGTCTGGGGGAGCCAGCCAGTGGGGGTGGAAGAGCCGTGA 13737
 1333 sNValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeu 1399
 13738 ATGTGGGACACCGGCGAGAGCTCCCACTTCTGGGCTATCCGTATGC 13787
 1400 ArgAsnProThrAspArgSerLeuLysGlyLysLeuMetAsnSerLys 1416
 13788 AGAATCCAAACCAACATCTTAAAGGCAACTCATAGAGAGAGAA 13837
 1416 sLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerPro 1433
 13838 ACTGCTCTACTACTGCTTCAAAACGAGGCTTCAATCCCGGAGG 13887
 1433 LeuLeuGlnProGlyLysThrAlaLeuAlaProLysLysArgSerArg 1449
 13888 CCTGCAACCTTGGGGGAGCTGCTGCGCTTAAAGAGAGGCGGAA 13937
 1450 GlyArgAlaGlyAla .HisGlyLeuSerLysGlyProLeuGlnLysArg 1466
 13938 GCGCGGGCAGGGGCGCCNATGACTCTCCAAAGCGCGCTGAGAGG 13987
 1466 roTyLeuGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGly 1482
 13988 CCAATCTTGGCCCGGCTCTCTCTGACTCCCGAGAGCGGGCCAGTGC 14037
 1483 ThrGlnGlyAlaSerGlnAspAsnSerGlyGlyGlyLysProLys 1499
 14038 ACACAAGGGCGGACGTGAGACAACTGTGTGAGAGGAGCAAGACCA 14087
 1499 sMetGlnGlnLeuGlyProAlaSerGlnProProGlnGlnArgProCys 1516
 14088 GATGGAGAGCTGGGCTGGCTCCCAACCCCGAGAGGCGGCGCTGGC 14137
 1516 InProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnLysSer 1532
 14138 AGCCCAAGCAAGGAGCAAGAAACAGCCAGGCGACCACTACAGAGC 14187
 1533 TyrSerLysArgLysArgLeuThrArgGlyValArgAlaLysAsnThr 1549
 14188 TATTCAGCGGAGCGCTCACTCGGGGCGGGCGCAAGAACCACTCTC 14237

1549 rSerProCysLysGlyValArgAlaLysArgArgGlnGlnValLeu 1566
 14238 TCCACCTCTGAAGGGGCGTGGCAAGCAGCAGCAGCAGCAGAGTCTCC 14287
 1566 roLeuAspProAlaGluProGlnLysLeuLysThrLysSerCys 1582
 14288 CCTGTGATCCCGCAGAGCCTGAAATCCGCTCAATTAATTTCTCTTC 14337
 1583 LysArgLeuArgSerAspSerArgThrProAlaPheSerProPheVal 1599
 14338 AAGCGCTGAGTCAAGACCGAGACCCCGGCTTCTCAACCTTGGTGC 14387
 1599 gValGlnLysArgAlaPheThrThrLysThrValAlaAsnSerP 1616
 14388 GGTGAGAGAGCGAGACGCTTACACCATATCATCTGTCTCACTCC 14437
 1616 roLysAlaProLysProHisArgLysProSerSerAlaSerSer 1632
 14438 CTGGAGATGCGCCAGCCCAAGCCCAAGAGAGCTTCTCTGCTGCTCT 14487
 1633 SerSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeu 1649
 14488 TCCATATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14537
 1649 arThrLeuProGlyLysLysLysLeuGlnProArgProSerLeuPro 1666
 14598 CACACTCCCTGAGAGCTTCATCTGACGCGGCGGCTCTCTGCTCT 14587
 1666 rSerThrMetHisLeuGlyProValAlaSerLysAlaLeuSerThr 1682
 14588 CTTCCAGATGACACTTGGGGGCTGTGGTTTCCAAAGCCCTGAGACT 14637
 1683 CysLeuValCysCysLeuGlnAsnProAlaAsnPheLysAspLeu 1699
 14638 TGCCTGTGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14687
 1699 YAspLeuCysGlyProTyrProGlnHisCysLeuProLysLysP 1716
 14688 GAGCTCTGTGGGCTCTACTACCTGAACTCTCTCTCTCTCTCTCT 14737
 1716 roLysLeuLysGlnLysValArgProGlnGlyThrCysGlnAlaSer 1732
 14738 CAATCTCAAGAGAGAGTGTGGGCGAGAGACCTGAGAGAGAGCTG 14787
 1733 LeuProLeuGlnArgThrLeuLysGlyProGlnCysAlaAlaAlaThr 1749
 14788 CTCCGCTTGAAGAGAGTCTCAAGGTCCGAGTGTGAGCTGCGCAC 14837
 1749 rAlaGlyLysProArg 1755
 14838 TGGCGGGAAGCCCGCCAGG 14856
 seq_name: gp_pat:AX024201
 seq_documentation_block: 5395 bp DNA PAT 15-SEP-2000
 LOCUS AX024201
 DEFINITION Sequence 1 from Patent WO0008143.
 ACCESSION AX024201
 VERSION AX024201.1 GI:10184512
 KEYWORDS
 SOURCE human;
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 5395)
 Poustika, A., Lehrach, H., Radeflof, U. and Seranowski, P.
 Gene isolated on the short arm of human chromosome 17
 Patent: WO 0008143-A. 1 17-FEB-2000.
 DEUTSCHES KREBSFORSCH (DE); POSTSTA ANNMARIE (DE); LEHRACH HANS
 (DE); RADELLOF UWE (DE); SERANOWSKI PETER (DE); MAX PLANCK INST
 FUER MOLEKULAR (DE)
 FEATURES
 Location/Qualifiers

384 IAGIYIIEHRASPHISSEHSISPhMETProLEULEUENProSERPro 400
1085 CAGGACTACTGACACAGCCACTCATGCCCCCTGCTCAATCCCTCCCA 1134
401 THRSPALATHRSERSEVALASPTINGLALAGLYANCYSLYSProLE 417
1135 ACGGATGCCACAGCTCTCT 1155
417 UGLNLYASPLYSLAUProGLUASNLEUSERASPLEUSERLEUINS 434
1156 GCKG 1159
434 ERLEUTHRALAUHRLAUGLIVAGLJHNSNLESERASPTIRVALGIN 450
1160 ACCTCAGCGCTGACCTCACAGGTGAGAACATCTCCAAACCGCTCCAG 1209
451 GLNLEULEUSERLYSALALAVAlProGLNLYSLYSGLYVALLYSAS 467
1210 CAGCTGCTGCTCCAAAGGCTGCTGTGCCGCAAGAAAGGTGCAAGAA 1259
467 NLEUVALSERATGTHProGLNGLNHSLSYSSERGLNHSYSSERProG 484
1260 CCMCTGTCCAGAGACCCAGACAGCATAAACCCAGACTGCAAGCCCG 1309
484 LUGLYSERGLYTSERLALAGLUPROALAGLYTHProLEUSERGLUPro 500
1310 AAGGAGCGGCTACTACGCCGAGGCCGAGGACACCGCTGCAAGACCG 1359
501 PROSERSETHRProGLNINSETHRHSALAGLUPROGLNVALASPTY 517
1360 CCGAGCAGCAGGACAGTCCAGCGATCGGAGCCGCAAGAGCGGCTA 1409
517 RLEUSERGLYSERGLUASPPROLEUGLUNAGYSERPHELEUTHRYCYSAN 534
1410 CCTGAGCGGCTCCAGAGACCACTGAGCGAGCTTCCTCACTGCAACC 1459
534 LNALARGLYSERPROALARGVALASNSEASNSERYSLALYSPro 550
1460 AGGCGCGTGAGGCCCTGCGAGGCTCAACAGCAACTCGAAGGCCAAGCC 1509
551 GLUSERVALSERTHRCYSERVALTHRSERPROASPSMETSETHLY 567
1510 GACTCGGTGTCCACTGTCTGTGACCTCTCTGACACATCTCCACCA 1559
567 SSERASPSSERPHEGLINSETHRHSGLYSERLEUPROLEUASPSER 584
1580 ATCTGAGCACTCTCCAGAGCTTACAGGAGTCTGCTGCGCTGACACT 1609
584 HESERTYSPHEVALALAGLYLUAARGSPYSProARGLLEULEUSER 600
1610 TCTCCAGTCTGCGCGGTGAGCGGAGTGTCCGCGCTGCTGCTCAGC 1659
601 ALALEUALAGLNUASPLEUALASERGLULEUGLYLEUGLNUAL 617
1660 GCCCTGGCAGAGAGACCTGCGCTCCAGATCTCTGGGCTGCAAGAGC 1709
617 ALLEGLYLULYSALASPLYSLATTPALAGLUALProSERLEUVAL 634
1710 CATCGGTGAGAAAGGCCGCAAAAGCTTGCGCTAAACACCGCTGCTGA 1759
634 YSASPSSETHRCYSERVALTHRSERLEUGLUNHSISERLACYSLAU 650
1760 AGGACAGCAGCAAGCCCTCTCTCGTGGAGAACACACCGCTGCTG 1809
651 ASSEVALALALYSERLALATTPROARGProGLYGLUProGLUALALE 667
1810 GACTCTGTGCGCAAGAGTGTGCGCCCGCTGAGGAGCGGAGCCCT 1859
667 UPROASPSERLEUGLNUASPLYSLYGLYASNALALYSASPSSETH 684
1860 GCCGACTCTCTGCACTGAGAGGCGGCAATGCTCAAGACTCTGACG 1909
684 ROGLYLEUPHEGLUNASPPROSERVALALAPHEALATHProASPPROLYS 700

1910 CAGGCTGTGTGAAGACCTTCCTGGGCTTCTGCTGACGCTGACCCAAA 1959
701 LYSTHRTINGLYPROLEUSERPHEGLYTHLYSPROTHLEUGLYVALP 717
1960 AAGACACTGTGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2009
717 CALAPROASPPROTHRALALAPHEASPCYSPPHEPROASPHRTHTA 734
2010 TGTCCAGAACCCCACTACAGCACTTTGATGCTTTCCGAGCAACCG 2059
734 LALASERSETHRALASPSERLALASPPROALATTPROGLUASN 750
2060 CTGCGAGCTCAGCGGACAGCGGACACCCCTTGTGCTGCGCAAGGAAAAC 2209
751 LEUGLYASPALACYSPROARGTGTGTYLEUHSISPROGLYGLUETHLY 767
2110 CTGGGGGAGCTGTGCCAGGTGGGATTTGACCTGCGAGACTTCCAA 2159
767 SGILEUGLNUGLNGLYSLASALASERTASGLYLESERTYSGLYASPT 784
2160 GGGCTGAGCAGGAGTGGAGAGGCTCAGATGSCATCAGCAAGGGACA 2209
784 HRHSGLUALASERLACYSLAUGLYPHEGLNGLNUASPPROGLY 800
2210 CCGATGAGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2259
801 GLULYSVALALASERLEUPROGLYASPHYSLYGLNUALVALGLYL 817
2260 GAGAGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2309
817 YVALYSGLUUALAGLYGLYLEULEUINCYSPROGLVALALATYSA 834
2310 GGTGAGAGAGAGGAGGTGGCTGCTGAGAGGCTGAGAGGAGGAGG 2359
834 LAAPARGTPLEUGLUNASPSERARGHSISCYSETHRALASAPHE 850
2360 CTGACCGGTGCTGAGAGACGCGGACTGCTGTTCACCGCGACTTC 2409
851 GLYASPLEUPROLEUPROTHRSERARGYSLYGLNUAL 867
2410 GGGAGCTCCACTGCTGCAACCCAGAGAGGAGGAGGAGGAGGAGG 2459
867 AGLNGLUTYRSETHRCYSERGLULEULEUINLYSERPROGLNGLA 884
2460 TTAGAGAGAGTACTCTCTCTATGTGAGCTCTGAGGAGGAGGAGG 2509
884 RPROGLYMETGLUNASPPROLEUSERPROLYSALAPROLEULIACYSTR 900
2510 GGCCTGGATGAGGAGCCGCTGTACCCAGGAGGAGGAGGAGGAGG 2559
901 LYSGLUGLNUALAGLNUALNUASPSERLYSALAGLYTPGLYSETP 917
2560 AAGGAGAGGTGGAGAGGTGTGTGACTCCAAAGCGGCTGCTGCTC 2609
917 OCYSHISLEUSERGLYLSERVALILEULEUGLYPROTHRALISLYT 934
2610 GTGCCACTCTCAGGGAGTCTCTGCTACCTGCTGCTGCTGCTGCTG 2659
934 HGLNLSERLYSVALGLINSETPPHEGLINSETHRSERHSISMETLY 950
2660 CCGAGTCAAAGGTCCAGACTGTGTGAGTCTCTGTGCAACATAGAG 2709
951 PROGLYGLUGLNUGLYPROASPLYGLNUARGLAPROGLYASPSERTHT 967
2710 CCGAGTGAAGAGGCTGTGTGAGGAGGAGGCTCAGGGAGATTCACAC 2759
967 RSEASPSALASERLEUALAGLNUASPSERLYSVALAPROGLUAL 984
2760 CTGAGAGGCTCTGCTGCGCAGAGACCCAGAACCTGCTGCTGCTGAG 2809
984 LARPROLALALYSGLUProVALProAGGLYSETHRSERLEUARGSER 1000

2810 GCGCCATCGCAAGAGAGCTGTGCCAGGGGCAAAAGCTTACGGAGC 2859
1001 ArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017
2860 CGTCGGGGTGCACGGGGGCTGCCAGGCGGAGGACTCCCATGCGAGGC 2909
1017 aProValLeuProLysAspLeuLeuProGluSerCysThrGlyProp 1034
2910 ACCATGCTGCCAAGAGACCTCTTGCTCCCTGATCTCTGACAGGGCCCC 2959
1034 roGlnGlyGlnMetGluGlyAlaGlyAlaProGlyArgGlyAlaSerGlu 1050
2960 CCAAGGAGAGATGGAAGGGCTGAGAGCCCGCCAGGCCGGGGCTCGGAA 3009
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067
3010 GGGCTCCCGCAGGATGTACTCTGCTCTCACGGCCCTGAGTGGCCCG 3059
1067 gThrProGlyProGlyLeuThrThrProAlaProProAspLysL 1084
3060 CAGGCCCGGACCCCGAGGCTGACACACCCCTGCACCCCGCCAGCAAC 3109
1084 euGlyGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100
3110 TGGGGGCAAGCAGGAGCCCTTCAAGTCGGGCAAGCGGTGGGGAAG 3159
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117
3160 CCGTCACCAAGCTGCCCTCCAGCCCGCAGCAACCCCGCGCTGCCGT 3209
1117 AlaSerAspSerProMetGlySerLysThrLysThrLysThrAspSerP 1134
3210 GCGCTCCGACAGCGCCGCTGAGGGCTCCAGGACCAAGAGACAGACTAC 3259
1134 roSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThr 1150
3260 CCAGCAGCGCTGGCAAGGACGAGCGCTCCATCATCTTCGTCACGCACC 3309
1151 LysThrGlnGluIlePheHisSerLysArgArgArgProSerGluGlyAr 1167
3310 AAACCCAGAGATCTTCATCCAGCGCGGAGGCCCTCTGAGGGCGG 3359
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisLeup 1184
3360 GCTCCCAACTGCGCTGCCACCAAGAGCTCTCGACACAGCCACTTGC 3409
1184 roAlaThrPheLysValSerSerProGlnLysGluGlyArgValSer 1200
3410 CCGCCCATTCAGGTCTCCAGCAGCCCGCCAGAGAGGCGCAGGTGAGC 3459
1201 GlnArgAlaArgValProLysProGlyValGlySerLysLeuSerAspAr 1217
3460 CAGCGGGCAAGGTCCCAAACTGTGTGCGAGCGCAAGCTCTTGACCG 3509
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValProt 1234
3510 GCGCTCCAGTCCGCTCAAGAGAGTGGCCCTTCATGGCGCGGTCCCA 3559
1234 hrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerAsn 1250
3560 CCAAGAGCGGAACTGTCTTCCGAGCGCCGACACAGCAGCAGCAGCAAC 3609
1251 AlaSerAlaMetGlyGluMetGlyArgArgGlyLeuArgValProp 1267
3610 GCGAGTGGCAATGGGGAGATGGGAGGAGGAGGAGGCGCTGAGGGTTCCC 3659
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProProArg 1283
3660 CACCTCTTCAAGAGGATGTCTTCTCCCAAGAAAGCCAGCCACCCAGG 3709
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProProPhe.L 1300
3710 GCATGGCGGAGCTGCCAAGAGCTCCCAACCCCGGAGACCCCGCATGCC 3759

1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLys 1316
3760 TGCTTCAAGCTCGCTCTCGGCAGCCTTCCAGGGGGCCCAAGAGACCA 3809
1316 sValLeuProProArgLysGlyArgGlyLeuLysLeuGluAlaIleValG 1333
3810 GGTGCTGGCCACCCCGAAGGGCGGGGCTGAAGCTGGAAAGCCATCGTGC 3859
1333 LnyLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349
3860 AGAAGATCACTCCCGCCAGCCTCAGAGAGTCCCATCTAAGAGCCAGGG 3909
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgG 1366
3910 GCTCTCTCTGTGTAATCTCTGAGCCCATCTTCCGAAAGACCGGTGG 3959
1366 YLeuLysGlyAlaGlyLysProValGlyValGlnGlyLeuVala 1383
3960 GCTCAAGGGTCTGCGGGGAGCCAGTGGGGGTGGAGAGGCCCTGTGTAA 4009
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCys 1399
4010 ATGTGGGACCGGGCAGAAAGCTCCAACTTCTGGGCTCATCCGTTATOC 4059
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysL 1416
4060 ASNAATCCACCCACAGATCTTTAAAGGCAAACTCATGAAAC..... 4101
1416 sLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGlu 1433
4101 4101
1433 laLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449
4101 4101
1450 GlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgPr 1466
4101 4101
1466 oTyrLeuGlyProAlaLeuLeuThrProArgAspArgAlaSerGlyT 1483
4102AGGCCAGTGGCA 4114
1483 hrGlnGlyAlaSerGluAspAsnSerGlyGlyLysLysProLys 1499
4115 CACAAGGGCCAGTGGAGGACAACTCTGTGGAGGAGGCAAGAGCCAAAG 4164
1500 MetGluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysG 1516
4165 ATGGAGGAGCTGGGCTGCTCCCGCCCGGGAGGCGAGGCCCTTGCCA 4214
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSert 1533
4215 GCCCAGACAAAGGGCAGACAGAAACAGCCAGCCACCACTACAGCAGCT 4264
1533 yrSerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549
4265 ATTCCAGCGGAGCGCTCTACTCGGGGCGGGGCCAAGAACACCACTCT 4314
1550 SerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuPr 1566
4315 TCACCTCTTAAGGGGCTGCCAAGCAGCAGCAGCAGCAGCAGGTGTGCC 4364
1566 oLeuAspProAlaGluProGluIleArgLeuLysTyrIleSerSerCysL 1583
4365 CTGTGATCCCGAGAGCCTTGAAATCCGCTCAAGTACATTTCTCTTGCA 4414
1583 ysArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599
4415 AGCGGCTGAGGTGAGAGCGGAGCCCGGCTTCTCACCTTCGTGGGG 4464

684 roglyLeuPheGluAspProSerValAlaPheAlaThrProAspProlys 700
1910 CAGGGCTGTTGAAAGACCTTCGCGGCTTCGTACGCTGAGCCCAA 1959
701 LysThrThrGlyProLeuSerPheGlyThrIysProThrLeuGlyValPr 717
1960 AAGACACTGTCCTCTCTCTCTTGTGTACCAAGCCACCTTGGGTTCC 2009
717 AlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThr 734
2010 TGCTCCAGACCCCACTACAGACCTTTGACTGTTCCCGAGACACACCG 2059
734 AlaAspSerAlaAspSerAlaAspProPheAlaThrProGlyValAsn 750
2060 CTCACAGCTCAGCGAGACGCGCCACCCCTTGGCTGCGAGGGAAC 2109
751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyLeuThrIy 767
2110 CTGGGGATGCTTGTCCAGATGGGATGGACCTCGCGAGCTTACAA 2159
767 sGlyLeuGluGlnGlyGlyLysAlaSerAspGlyIleSerGlyAspT 784
2160 GGCCCTGGAGCAGGGGGGAGGAGGCTCAGATGATCAGCAAGAGGAG 2209
784 hHisGluAlaSerAlaCysLeuGlyPheGlnGluAspProGly 800
2210 CCATAGAGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2259
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGly 817
2260 GAGAGGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2309
817 yValLysGlnGluAlaGlyGlyLeuLeuGlnCysProGlyValAlaLys 834
2310 GGTGAAG 2359
834 LAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850
2360 CTGACCGGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2409
851 GlyAspLeuProLeuLeuProThrSerArgLysGluAspLeuGluAl 867
2410 GGGAGCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2459
867 agLingLingLysSerSerLeuGlnLeuLeuGlySerProGluGln 884
2460 TGAGAGAGAGTACTCTCTCTGATGAGCTCTGAGAGAGAGAGAGAG 2509
884 rProGlyMetGlnAspProLeuSerProLysAlaProLeuIleCysThr 900
2510 GGCTGTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2559
901 LysGlnGluValGlnGluValLeuAspSerLysAlaGlyTrpGlySerPr 917
2560 AAG 2609
917 oCysHisLeuSerGlyLysSerValIleLeuLeuGlyProThrValGlyT 934
2610 GTGCCACTCTCAGAGAGAGTCCGTATCTGCTGAGAGAGAGAGAG 2659
934 hrgLysLeuLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950
2660 CCGAGTCAAGAGTCCAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2709
951 ProGlyGluGluGlyProAspGlyGluArgAlaProGlyAspSerThrTh 967
2710 CCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2759
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGlu 984
2760 CTCGAGAGCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2809
984 LAspThrLeuAlaLysGluProValProArgGlyLysSerLeuArgSer 1000
2810 CCGCATCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2859
1001 ArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017
2860 CATTGGGTGACCGGGGGGTGGCGGAGCGGAGAGAGAGAGAGAGAGAG 2909
1017 aProValLeuProLysAspLeuLeuLeuProGluSerCysThrGlyPro 1034
2910 ACCAGTGTGTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2959
1034 roGlnGlyGlnMetGluGlyAlaGlyAlaProLysArgGlyValSerGlu 1050
2960 CCGAGGAG 3009
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067
3010 GGCTTCGCCAGAGATGTACTGTTCTCTCTCTCTCTCTCTCTCTCT 3059
1067 gThrProGlyProProGlyLeuThrThrThrProAlaProAspLysL 1084
3060 CAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3109
1084 euGlyGlyLysGluArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100
3110 TGGGGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3159
1101 ProSerProLysAlaAlaSerSerProSerAspProAlaProValPro 1117
3160 CCTTCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3209
1117 LAlaSerAspSerSerProMetGlySerLysThrLysGluThrAspSer 1134
3210 GGCTTCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3259
1134 roSerThrProGlyLysAspGluArgSerMetIleLeuArgSerArgThr 1150
3260 CCAAG 3309
1151 LysThrGlnGluIlePheHisSerLysArgArgArgProSerGluArg 1167
3310 AAAAGAGAGAGATCTTCACTCAAGAGAGAGAGAGAGAGAGAGAGAG 3359
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspHisSerHisLeu 1184
3360 GCTCCCACTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3409
1184 roAlaThrPheLysValSerSerSerProGluLysGluGlyArgValSer 1200
3410 CCGCAGATTCAGAGTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3459
1201 GlnArgAlaArgValProLysProGlyValGlySerLysLeuSerAsp 1217
3460 CAGCGGAGAGAGTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3509
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValPro 1234
3510 GCCCTCATGAGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3559
1234 hLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerAsn 1250
3560 CCAAG 3609
1251 AlaSerAlaMetGlyGluMetGlyArgArgArgArgGlyLeuArgValPro 1267
3610 GCCATGGCAATGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3659
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProProArg 1283
3660 CACCTCTTCAAGAGAGATGTTCTTCCCAAGAGAGAGAGAGAGAGAGAG 3709
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProPheL 1300


```
3710 CCAATGGCGAGCGCTGCCACAAGCTCCCAAGCCCGCGGAGACCCCGCATGCC 3759
1300 euProGlnAlaArgLeuSer. AlaAlaPheGlnGlyAlaMetLysThrLys 1316
|||||
3760 TGCCCTCAAGCTCGGCTCTCGGGCAGCGCTTCAGGGGGGCCATGAAGACCA 3809
1316 sValLeuProProArgLysGlyArgGlyLeuLysLeuGluAlaIleValG 1333
|||||
3810 GGTGCTGCCACCGCCCGGAAGCGCGGGGCTGAAGCTGGAAAGCCATCGTGC 3859
1333 InLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349
3860 AGAAGATCACTCGCCAGCGCTCAAGAGTTCGCATGTAAGCGCCAGGG 3909
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysAspArgL 1366
|||||
3910 GCCTCTCTGTAATCTCTGAGCCCATCCCTTTTCGACAAAGACCGTGG 3959
1366 yLeuLysGlyAlaGlySerProValGlyValGluGluGlyLeuValA 1383
|||||
3960 GCTCAAGGGTGTGGGGCAGCCAGCTGGGGGTGGAGAGGCGCTGGTAA 4009
1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCys 1399
4010 ATGTGGGCACCGGGCAGAGCTCCAACTTCTGGGGCTGATCGGTTATGC 4059
1400 ArgAsnProThrAspSerLysLysGlyLysLeuMetAsnSerLysLys 1416
|||||
4060 AGAAATCCAAACCAAGATCTTAAAGGCAACATCATGAAC 4101
1416 sLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGluA 1433
4101 ..... 4101
1433 laLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449
4101 ..... 4101
1450 GlyArgAlaGlyAlaHisClyLeuSerLysGlyProLeuGluLysArgPr 1466
4101 ..... 4101
1466 ofyLeuGlyProAlaLeuLeuThrProArgaspArgAlaSerGlyT 1483
4102 ..... AGGCGCAGTGGCA 4114
1483 hrGlnGlyAlaSerGluAspAsnSerGlyGlyLysLysProLys 1499
4115 CACAAGGGCCAGTGGAGCACTCTGTGGAGGAGGCAAGAGCAAG 4164
1500 MetGluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysG 1516
4165 ATGGAGGAGCTGGGCTGGCTCCAGCCCCCGGAGCGAGGCCCTGCCA 4214
1516 nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerT 1533
4215 GCCCCAGACAGGCGCAGAAACACCCAGGCCACACACACACACACAC 4264
1533 yfSerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549
4265 ATTCCAAGCGGAGCGCTCACTCGGGGCGCGGCGGCAAGAACACACCT 4314
1550 SerProCysLysGlyArgAlaLysArgArgGlnGlnGlnValLeuPr 1566
4315 TCACCTGTGAAGGGCGTGCACAGGCGGAGGAGGAGGAGGAGGAGGAG 4364
1566 oLeuAspProAlaGluProGluIleArgLeuLysTyThrIleSerSerCys 1583
4365 CCTGGATCCCGCAGAGCTGAATCCGCTCAATGATCTTCTCTTCCA 4414
1583 ysArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599
4415 AGCGGCTGAGGTCACAGACGGGACCCCGGCTTCTCACCTCTGTGGGG 4464
```

```
1600 ValGluLysArgAspAlaPheThrThrIleCysThrValValAsnSerPr 1616
|||||
4465 GTGGAGAGCGAGAGCGCTTCAACCATATGCACTGTGTCAACTCCCC 4514
1616 oGlyAspAlaProLysProHisArgLysProSerSerSerAlaSerSerS 1633
4515 TGGAGATGCGCCCAAGCCCAAGAGAGGCTTCTCTCTGCTGCTCTCTT 4564
1633 erSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAla 1649
4565 COTATCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4614
1650 ThrLeuProGlyLysIleLeuGlnProArgProSerLeuProLeuSe 1666
4615 ACACCTCCCTGGAGGCTCCATCTCGACGCGCGGCTCTCTCTCTCTCT 4664
1666 rserThrMetHisLeuGlyProValValSerLysLysAlaLeuSerThr 1683
4665 CTCACGATGCACTTGGGCTGTGGTTTCCAAAGGCTCTGAGTACCTCT 4714
1683 ysLeuValCysCysLeuCysGlnAsnProAlaAsnPheLysAspLeuGly 1699
4715 GCCTTGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4764
1700 AspLeuCysGlyProTyrrProGluHisCysLeuProLysLysLysPr 1716
4765 GACCTCTGTGGGCTACTACCTGACACTGCTCTCTCTCTCTCTCTCTCT 4814
1716 oLysLeuLysGluLysValArgProGluGlyThrCysGluGluAlaSerL 1733
4815 ARAACTCAAGAGAGGTGGCGGCAAGAGGCACTGTGTGAGGAGGCTCGC 4864
1733 euProLeuGluArgThrLeuLysGlyProGluCysAlaAlaAlaThr 1749
4865 TGCGCTGTGAGAGACACTCAAGGTCCCGAGTGTGAGTGTGAGTGTGAG 4914
1750 AlaGlyLysProProArg 1755
4915 GCGGGAGAGCCCGCCAGG 4932
seq_name: gb_pr:AB058723
seq_documentation_block:
LOCUS AB058723 5915 bp mRNA PRI 05-JUN-2001
DEFINITION Homo sapiens mRNA for KIAA1820 protein, partial cds.
ACCESSION AB058723
VERSION AB058723.1 GI:14017856
KEYWORDS
SOURCE:
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites) Homo sapiens brain cDNA to mRNA, clone:hh01321.
AUTHORS Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 8 (2), 85-95 (2001)
MEDLINE 21245130
REFERENCE
2 (bases 1 to 5915)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdna1info@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
FEATURES
location/Qualifiers
source 1..5915
/organism="Homo sapiens"
/gb_xref="taxon:9606"
/clone="hh01321"
```


/tissue_type="brain"
/note="vector:pb1uescriptII SK plus"
458..5392
/gene="K18A1820"
CDS
/gene="K18A1820"
/note="Start codon is not identified."
/codon_start=1
/product="K18A1820 protein"
/protein_id="BAB47449.1"
/db_xref="GI:14017857"

translating="PARMOSFERCGRFHGKQNTYQTSERILENTRPSQAGLSC
DROBLADKYNPCEYSEYAGATPSGAAVADYHGSALPELQGLQGPAP
GYGVDSPYGPGRAGEESLOANGAPQPPQPLPAGVAKIDEMKMGTAHPSPO
YAEQAGQVPRPHSLHNOCPPOPLAPKLOHNDIAEPFPGHPRQHSQ
SEPTSSYSSVQGGGASHYSCTAPRPHORPLTASSSIAPGQRYOMLHAYOG
RLSYDQOQOQOQOQOQOQALQSRHHAQETLHONLAKYHOCQOQOQOCPDPAVE
PEQYCTPSSSHSPANSYGRSPYSSTPSLMELENLEPYSQPLSTGAPAGITD
HSHMPLNPSPTDASSVDQAGNCKPLQKDKLEMLDLSLQSLTASQVENS
NTYQOILLKKAAPQKGVKNLYSRTPQKQSOHCSPGSGSYAEPAATPLSEPSST
POSTHAEPOADYLSGSEDPLEPSFLYCNQANGSPARNSAKESPESVTSYSD
DMSTRKSDSPQLHGLPLDSEKFTVAGRDOPRLLLSALQDLASRLLQIOLARIGR
KADRAMAASIVKDSKPPPLENHSACLDVANSAMPREPRLPSLQIDKNGN
AKDSPLLEDESVAFATPDKKTGPTLGFGRKPTLVAPADPTAATPCPEDTLAS
SADSNFPAPEENLGDACPRMGLHPELTKLEQGGASDGLSKGDIETASACLGQV
BEDPGRKIASLPGDFKQEVGYKEEAGGLLOCEVAKADRMLEDRCCESTADFD
LPPLPTSKEDILABEETSSLCILDSPPQGRQMDPLSKAPILCTKEVEEYLS
KAGSGPCHLSSESVLLGPTVGTESVQVSEBESSMKRGEGRGPRGARAQSTTS
DASLAQKPNKPAVEPAJLAKREPVRGRKSLSRVRHGLPAEDSPORAPVLPKILL
PECTPGPOGMEGAPAGRGASGLPRMOTSLTALSEPTEPQPELTTTAPDKL
GGKQRAFKSGKEVGPSPKASSPSNPLPAVADSSPMGSKTETDPSPPGDKR
SWILRSTKQETEPHSHRRRPSRGLPNCRAKRLINDSHLPATFYSSSPKEGRYS
QARVPKGAQSKLSRPLHALKREKSAFMAVPPTKRNLYLSRSSSSNAGNGGDC
KEEREPSPLEFRKSSPKAKPTKNGEPAKRLPPTPPACCKLASRAAOGAMKT
KVLPRKRGRLKLEALYOKLTPSPILKFAKRAAGSAPNPISPLSKDRGKKGGS
PYVGEGLVYVGTOKLPTSGADPLCRNPNLSLKGKLNKSLSTDCRKEATSP
EALQGGTALAPKKRSKRGAGAGLSKGLKRYPLGALLLPDRASGTOGASED
NKGSGGKPKMEELGLASOPRGRPOCORAKOGPHNYSYKRRKTRAKNT
TSSPCGAKARRKQOQVLEPDPAPEIRIKYVTSCKRLKRSRTPARSPLEBLSAON
PRSGQNPAPADARPLCTTRDRNRYSGRGR"

BASE COUNT 1255 a 2038 c 1746 g 876 t
ORIGIN

alignment_scores:
Quality: 8355.50 Length: 1741
Ratio: 5.092 Gaps: 8
Percent Similarity: 94.256 Percent Identity: 92.648

alignment_block:
US-09-508-821b-6 x AB058723

Align seg 1/1 to: AB058723 from: 1 to: 5915

1 MetGlnSerPheArgGlnArgGlyPheHisGlyLysGlnInlInlAsnTy 17
|||||
470 ATGCAAGCTTTTCGAGAAAGGTGGTTTCATGCGAAACAGAGAACTA 519
17 TGINlInlSerGlnGlnlInlSerArgLeuGlnInlAsnTyArgGlnPro 34
|||||
520 CCAGCAGACCTCGCAGAAATCATCAGCTAGAAATTAAGCAGCAGCA 569
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTy 50
|||||
570 GTCAGGCGCGGCTAAGCTCGACCGGCGCTGCTCGCAAGGAGCTAT 619
51 TyrAsnProGlnProTyProSerTyGlnGlnGlyAlaGlnlInlPro 67
|||||
620 TATAACCCGACGCTTACCCGAGCTATGAGGTGCGCTGGCAGCGCC 669
67 TGLTThrAlaAlaAlaValAlaAlaAspLysTyrHisArgGlnSerLys 84
|||||
670 TGGCACTGACAGCGCGGTGCGCGCCGACAAATACACAGGAGAGCA 719

84 lAlaLeuProThlGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
|||||
720 CCCTCCACACACAGCAGAGCGCTGCAAGGGAGGGCGGCTTCCCTGGCTAC 769
101 GilyValGlnAspSerSerProTyProGlyArgTyAlaGlyGlnGlnGln 117
|||||
770 GGGCTCAGAGACAGACAGCCCTTACCCAGGCGGTATGCTGGTGGAGAG 819
117 rLeuGlnAlaTrpGlyAlaProGlnProProProProGlnProGlnPro 134
|||||
820 CCTTACAGGCTTGGGGGGCCCGACAGCCACCCAGCCAGCCGAGCCAC 869
134 euProAlaGlyValAlaLysTyrAspGlnAsnLeuMetLysTyrThrAla 150
|||||
870 TACCTCAGAGGGGTGCCAGATGATGAGAACTTGAGTAAAGACAGCA 919
151 ValProProSerArgGlnlInlTyAlaGlnGlnGlnGlnGlnGlnGln 167
|||||
920 GTGCCCCCAGCAGCAGCTTGCAGAGAGGGCCCGCAGGTGCTTCC 969
167 GThrHisSerLeuHisValGlnGlnProProProProGlnGlnProLeu 184
|||||
970 GACTCATCTCCGACAGCTCCAGCAGCAGCCGCGCCAGCAGCCCTGG 1019
184 lArgTyProLysLeuGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 200
|||||
1020 CATACCCCAAGCTCCAAAGGACAGAACTGCAAGACGATGCTGCTCCCT 1069
201 LeuProPheProGlnGlnlInlHisPheProGlnHisSerGlnSerPhe 217
|||||
1070 CTGCGCTTCCCGAGGAGTACCACTTCTCGACATTCCTCCAGTCTCC 1119
217 OThrSerSerThrTySerSerSerValGlnGlnGlnGlnGlnGlnGln 234
|||||
1120 CACTCTCTCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1169
234 lSerTyTyLysSerCysThrAlaProThlAlaGlnProHisAspArgPro 250
|||||
1170 ACTCCTTAAAGATTGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1219
251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeu 267
|||||
1220 CTGACTCGACGCTCGAGCTGCGCCCGGGGAGCCGGTCCAGAACTCTCA 1269
267 sAlaTyGlnSerGlyArgLeuSerTyAspGlnGlnGlnGlnGlnGlnG 284
|||||
1270 TCCCTACAGTCCGGCGCGCTCAGCTATGACAGCAGCAGCAGCAGCAG 1319
284 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
|||||
1320 AGCAGCAGCAGCAGCAGCAGCAGCAGCCTTCAAGAGCCGACACATG 1369
301 GlnThrLeuHisTyGlnAsnLeuAlaLysTyGlnHisTyGlnGlnGln 317
|||||
1370 GAACCTTCATTAACAAACCTCGCCAAATATACACACTACGGGAGCA 1419
317 nGlyGlnGlnTyCysGlnProAspAlaAlaValArgThrProGlnGln 334
|||||
1420 AAGCAGAGGCTACTGCGACCGCAGCAGCAGCGCGTGGAGCCCAAGAG 1469
334 yTyTyGlnThrPheSerProSerSerSerHisSerProAlaTrpSerVal 350
|||||
1470 ACTACAGACCTTACAGCCCACTCCAGCCACTCACCCCGCCGCTCG 1519
351 GlyArgSerProSerTyTySerSerThrProSerProLeuMetProAsn 367
|||||
1520 GGGCGCTCACCTTCTACAGTTCCACACCGGCGCGCTATATGCTAACT 1569
367 uGlnAsnPheProTySerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 384
|||||
1570 GAGGAACCTTCCCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1619

384 laGlyIleThrAspHisSerHisPheMetProLeuIleAsnProSerPro 400
1620 CAGGATCACTACACAGACCTTCATGCCCTTCCTCAATCCCTCCCA 1669
401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysProLe 417
1670 ACGGATGCCACCACTGTGTGGACACCCAGGCTGGCAACTGCAAGCCCT 1719
417 uGlnLysAspLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGlnS 434
1720 TCAGAGGACAGACCTCCCTGAGAACCTGCTGGATCTCAGCCATGCGAGA 1769
434 erLeuThrAlaLeuThrLeuGlnValGluAsnLysSerAsnThrValGln 450
1770 GCTTCAGCGGCTGACCTACAGGTGGAGAACATCCACACACGCTCAG 1819
451 GlnLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLysAs 467
1820 CAGCTGCTCTCCAAAGGCTGTGGCCGAGAGAAAGGTGTCAAGAA 1869
467 nLeuValSerArgThrProGluGlnHisLysSerGlnHisCysSerProG 484
1870 CCTGCTCCAGACCCACAGGACAGATAAAGCCAGCACTGCAGCCCG 1919
484 uGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSerGluPro 500
1920 AAGGAGCGGCTACTACAGCGAGCCGCGAGCACACCGCTGTCAAGCGG 1969
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGluAlaAspTy 517
1970 CGAGCAGACGCCACAGCTCCAGCATCGGAGCGGAGCGCGACTA 2019
517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCysAsnG 534
2020 CTTGAGCGGCTCCGAGACCCCTGAGCGAGGCTTCTTACTCAACC 2069
534 lnaAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550
2070 AGGCGGCTGGAGCCCTGCGAGGCTCAACACAACTCGAGGCCAAGGCC 2119
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567
2120 GAGTCCCTGTCCACCTGTCTGTGACCTCTCTGACACATGTCACCAA 2169
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584
2170 AFTGAGCACTCTTCCAGAGCTTACAGCGAGTCTCCCTCGACACT 2219
584 heSerLysPheValAlaGlyLysArgAspCysProArgLeuLeuSer 600
2220 TCTCAAGTTCGTGGCGGTGAGCGGACTGTCCGCGCTGTCTCAGC 2269
601 AlaLeuAlaGlnGluAspLeuAlaSerGluIleLeuGlyLeuGlnGluAl 617
2270 GCTTGGCTGACAGAGGACTGGGCTCCGAGATCTTGGGCTGCAGGAAGC 2319
617 alileGlyLysAlaAspLysAlaThrAlaGluAlaProSerLeuValL 634
2320 CATGGTGGAGAGGCGGACAAAGCTGGGCTGAAGACAGCCAGCTGTGTA 2369
634 ysAspSerSerLysProPheSerLeuGluAsnHisSerAlaCysLeu 650
2370 AGGACAGCAGCAAGCCACCTTCTGCTGGAGAACCAACAGCGCTGCTG 2419
651 AspSerValAlaLysSerAlaTTPProArgProGlyGluProGluAlaLe 667
2420 GACTCTGGCCAGAGTGGTGGCCCGGCTGGGAGCGCAGAGGCCCT 2469
667 uProAspSerLeuGlnLeuAspLysGlyLysAlaLysAspPheSerP 684
2470 GCCGAGCTCTTGCAGCTGGACAGAGCGGCAATGCCAAGGACTTCAGCC 2519
684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700

2520 CAGGCTGTGTGAAGACCTTCGTGGCTTCGCTACGCTGACCCCAAA 2569
701 DyeThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717
2570 AAGACAACTGTCTCTCTCTTGTGTAACCAAGCCACCTTGGGTTC 2619
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734
2620 TGTCCACACCCCTACACAGCTTTTGACTGTTCCTGCGGACACACCG 2669
734 laAlaSerSerAlaAspSerAlaAsnProPheAlaTTPProGluGluAsn 750
2670 CTGCCAGCTCAGCGAGAGCGCAACCTTTCCTGCGCAGAGAGAAAC 2719
751 LeuGlyAspAlaCysProArgTTPGlyLeuHisProGlyGluLeuThrLy 767
2720 CTGGGAGTGTTCCTCCAGGTGGGATTCACCTTGGCGAGCTTACCAA 2769
767 sGlyLeuGlnGlnGlyLysAlaSerAspGlyLysSerLysGlyAspT 784
2770 GGCTTGGAGAGGCTGGGAAGGCTCAGATGCATCAGCAAGGGGACA 2819
784 hrHisGluAlaSerAlaCysLeuGlyPheGlnGluAspProProGly 800
2820 CCTGAGGCTTCGGCTGCTGGCTTCCAGGAGGAGGACCCCTGCG 2869
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGlyG 817
2870 GAGAAGTGGCTCGTTCGCGGAGCTTCAAGCAGAGGAGGTGGTGG 2919
817 yValLysGlnGluAlaGlyLysLeuGlnCysProGluValAlaLysA 834
2920 GGTGAAGAGAGGAGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2969
834 laAspArgTTPLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850
2970 CTGACCGGTGCTGGAGGACGCGGACTGCTGCTGCTGCTGCTGCTG 3019
851 GlyAspLeuProLeuLeuProThrSerArgLysGluAspLeuGluAl 867
3020 GGGGACTCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3069
867 aGluGlnGluTyrSerSerLeuCysGluLeuLeuGlySerProGluGlnA 884
3070 TGAGGAGAGTACTCTCTCTATGTAGTCTCTGCTGCTGCTGCTGCTG 3119
884 rgProGlyMetGlnAspProLeuSerProLysAlaProLeuIleCysThr 900
3120 GGCTTGGATGAGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3169
901 LysGluGluValGlnGluValLeuAspSerLysAlaGlyTTPGlySerPr 917
3170 AAGGAGGAGTGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3219
917 oCysHisLeuSerGlyLysSerValIleLeuLeuGlyProThrValGlyT 934
3220 GTGCCACTCTCAGGGAGTTCCTGCTGCTGCTGCTGCTGCTGCTGCT 3269
934 hrGluSerLysValGlnSerTTPPheGluSerSerLeuSerHisMetLys 950
3270 CCGAGTCAAGTCTCAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 3319
951 ProGlyGlnGluGlyProAspGlyGluAlaProGlyAspSerThrTh 967
3320 CAGGTGAAGAGGCGCTGATGGGAGGAGGCTTCCAGSGGATTCACCA 3369
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984
3370 CTGGAGCGCTCTCTGCGCCAGAGCCCAAGCTGCTGCTGCTGCTGCTG 3419
984 laProIleAlaLysGluProValProArgGlyLysSerLeuArgSer 1000

1300	euproglnlaarglneuber	AlaaAaphegnlcyAlaMeluysThly	1316
4370	TGCTCGAAGTGGCTCGGGGAGGCTTCCAGGGGGCCATGAAGCCAA		4419
1316	swalleProProAarglysglyargglyleuylsleuglnAlaIleValG	1333	
4420	GGTGTCTGCCACCCCGAAGGGCGGGGGCTGAAAGTGAAGCCATCTGTGC	4469	
1333	InduyllethrsProserProserleuylsPheAlaCysLysAlaProgly	1349	
4470	AGAAAGACACCTCGGCCAGCTCCAGAGATTGGCATGAAAGGCCAGGG	4519	
1350	AlaSerProdlYasnProLeuSerProSerleuSerAspLysAspArgl	1366	
4520	GCCCTCTCTGGTAHCTCTGAGCCATCTCTTCCGAAAGACGGTGG	4569	
1366	YLeuylsglyAlaaglylyserProValglyValgluGlyleuVala	1383	
4570	GCTCAAGGGTGTGGGGGGACCCGACGGGGGGGAGAAAGAGCCGTGTA	4619	
1383	snValglyThrnglyInduylsleuProthrsSerGlyAlaAspProLeuLys	1399	
4620	ATGTGGGACCGGGGCGAAAGCTCCCAACTCTGGGGCTGAATCCCTHAGC	4669	
1400	ArgAsnProthrsAsnArgserLeuylsGlyLysleuMetAsnSerLysly	1416	
4670	AGAAATCCAAACAGACATCCCTAAAGGCAACCTAGAGACATAAGAA	4719	
1416	sleuSerSerthrsPcysPheLysThrGluAlaPheThrSerProGlu	1433	
4720	ACTGCTCTCTACTGACCTTTTAAACCGAGGGCTTCACATCCCGAGAG	4769	
1433	leuGlnProglglyThrAlaIleuAlaProLysLysAsnArgLys	1449	
4770	CCCTGAGCTGGGGGGAGTCCCTCTGGGCTTAAGAGAGAGCGGGAAA	4819	
1450	glyArgAlaGlyAlaHISglyLeuSerLysglyProleuGlnLysArgtr	1466	
4820	GGGCGGGGACGGGGCCCATGTGACTCCAAAGGCCCTCGAAGACGGCC	4869	
1466	oTryLeuGlyProAlaIleuLeuThrProArgAspArgAlaSerGlyT	1483	
4870	CTATCTTGGCCCGGGCTCTGCTCTCATACCCCGAGACAGGGCAATGGCA	4919	
1483	hrGlnLysAlaSerGlnAsnAsnSerGlyGlyGlyLysLysProLys	1499	
4920	CACAAAGGGCCAGTGAAGAAACTGTGGTGAAGAGGCAAGAACGCAAG	4969	
1500	MetGlnLLeuGlyProAlaSerGlnProProGlnLysArgProCysgl	1516	
4970	ATGAGAGACTGGGCTGGCTCCAGGCCCGGAGGGGAGGGCCCTGGCA	5019	
1516	nProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTrySerSerT	1533	
5020	GCCCAAGACAAAGGACAGAAACAGCCAGAGCCACCACTAAGAGAGCT	5069	
1533	YrSerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrser	1549	
5070	ATTCCAAGCGGAGAGGCTCATCTCGGGGGCGGGGCCAAAGAACACCACTCT	5119	
1550	SerProCysLysGlyArgAlaLysArgGlyArgGlnGlnValLeuPro	1566	
5120	TACCCCTGTAAGGGGCGTGGCAAGACAGACAGACAGAGAGGTGTGTGC	5169	
1566	oleuAspProAlaGlnProGlnIleArgLeuLysTrylleSerSerCysgl	1583	
5170	CGTGAGATCCCGAGAGAGCTGAATACGGCTCTCAAGTAATTCCTCTTGGCA	5219	
1583	ysArgLeuLysSerAsnSerArgThrProAlaPheSerProPheValArg	1599	
5220	ACGGAGCGAGCTTAAGACGGCGAGCCCGGCTCTCAACCCCTGGA...	5265	

1600 ValGluLysArgAspAlaPheThrThrLeuThrCysThrValValAsnSerPr 1616
5266GCCAGCC 5273
1616 oGlyAspAlaProLysProHisArgLysProSerSerSerAlaSerSers 1633
5274 TGGAGGCCA...AAACCAAGAAGCGGCCAGACGACCTCCGGCTCCG 5320
1633 erSerSerSerSer.....SerPheSerLeu 1641
5321 GCGAGCGCGGACCGCTGTGCGACACAGGAGCGCGCGCTACTCTGC 5370
1642 AspAlaAlaGlyAlaSerLeu..... 1648
5371 ACGGAGCAGGACAGCGCTAGATTTCGTACAAAACCTGTGTACCCCT 5420
1648 1648
5421 CTATATATATGTTACATAGATGATATATGTTGGGAACATGCTCGCTC 5470
1649AlaThrLeuProGlySerLleLeuGlnProArgPro 1661
5471 TCCCGTGTGCGCGCGCTGCGTGCCTGCGCGCGGACAGAGCCCGACCG 5520
1662 SerLeu..... 1663
5521 GCGCTTCCGCGGTAAAGGGCTACCGGAGCGGCACTTGTGCGAGCGCA 5570
1664ProLeuSerSerThrMethHisLeuGlyProValValS 1676
5571 CCAGCGCGCGCGCGCTGCGTGCCTGCGCGCGCTGCGCTGCGTGTGT 5620
1676 erLysAlaLeuSerThr 1681
5621 COTCAGCTCTGTCCAG 5637
seq name: gb_ro:MUSMA
seq_documentation_block: 7222 bp mRNA ROD 07-FEB-1999
LOCUS MUSMA
DEFINITION Mouse mRNA for unknown product, complete cds.
ACCESSION D29801
VERSION D29801.1 GI:475015
KEYWORDS ORF.
SOURCE Mus musculus postnatal-day1 brain cDNA to mRNA, clone_lib:library
of Stratagene (La Jolla, CA).
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Imai.Y.
TITLE Direct Submission
JOURNAL Submitted (04-APR-1994) to the DDBJ/EMBL/GenBank databases. Yuji
Tanabe; Osaka University Medical School, Molecular Neurobiology,
Tanabe; 2-2 Yamadaoka, Suita, Osaka 565, Japan (Tel:06-879-3646,
Fax:06-879-3648)
2 (bases 1 to 7222)
REFERENCE
AUTHORS Imai.Y., Suzuki.Y., Matsui.T., Tohyama.M., Watanaka.A. and Takagi.T.
TITLE Cloning of a retinoic acid-induced gene, Gtl, in the embryonal
carcinoma cell line P19: neuron-specific expression in the mouse
brain
JOURNAL Brain Res. Mol. Brain Res. 31 (1-2), 1-9 (1995)
MEDLINE 96078271
COMMENT Submitted (04-Apr-1994) to DDBJ by:
yuji Imai
Department of Molecular Neurobiology, Tanabe
Osaka University Medical School
2-2 Yamadaoka, Suita
Osaka 565
Japan
Phone: 06-879-3646
Fax: 06-879-3648
FEATURES
Location/Qualifiers

1. 7222
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_lib="library of Stratagene (La Jolla, CA)"
/dev_stage="postnatal-day1"
/tissue_type="brain"
309.5831
/codon_start=1
/product="Unknown"
/protein_id="BAA06184.1"
/db_xref="GI:475016"
/translation="MSPFRRCGPHGKQVQVPSQTSRLNYPQAGLSCDRQR
LLADYISQPVYGVGGTSPSTVATAADKVRHSGSKLQGRPAFYSYVQDSSYP
GRYSEEGTQGGGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
RTHSLYPP
VQGGGGAHSYKSPSAQPHRPMANLANAPQVQMLHAYQGRUGYEQQQAL
QGRHTEILHYNLAQHYQGGGQGYCPPTAVTTPYQYQYQYQYQYQYQYQYQY
RSPYSSTPSLPMLENFYQOPLSTGATFTTGHSHFMLLPSPSPSPSPSPSVG
QAGNCKLAKELIPNLLISYLSLTALTSOVENISNTVOQLLLSKATMPOKGVKN
LVSTPQHKSHQCSPEGSYSAEPAGTLPSEPPSPSPSPSPSPSPSPSPSPSP
ERSFLYQARGSPARVNSKAPSPVCSVTPSPDNDSTKSDSFQSLHSTLPLDS
FSKTVAGERCPLLLSALAQEDLASEILQEAIVEKADKAWAEASSLPKDNKPPF
SLENHGACLDIVAKTSWQGPETLPPELQDKGSKTDFSPGLFEDFSVAFATDP
KTSPLSEGTPLLTATPPTAAADCFDPTPTASSVDGANPFAMPENLGDACPR
WGLHPGLTGLGAKADGVGKADAHASACMGFOEDHAIKPAALSGDGFQOEA
EGVEVGGLLQCPVAKADONLESHHSCSSSTDFGLPLPPGKKELEAEERYSS
LOELGSPQPSLQDPSIAPLMCTKEAEALDTRAGWSPCHLSEPAVLIGPS
VGAQSKVSWFESSHMKPGEPEMERAPGSSSTGSLAPRPNKPAVPPEPTAK
BEVPRGSLRSRVRHGLPEAEADSPCRPALPKDLLPECTGPPQGAAGAPGRG
LSEGLPRMCTRSITALSPPOTPGPGLTTTPTPKLGGKQRAAFSGRRVGPSPKA
ASSPSNPAALPVASDSSPMGSKTEKDPSPMPKQORSWLRSRTKPPQVFAKRRP
SERIPDCRATKPLPANHLPTAKVSSGPKQEGMSQVVKPKGTGKGLSDRLHT
LKRKSAMFAPVAKESLILSNNGSGDGEREAPSPGLRRMASPQAPRPSGGE
PPPPPPPPAACMGILTSQSLSPAVRTKIPRGRGKLEAIVOKITSPGLKLAC
RVAGAPPTPRSPALPRPPGSPAGAEGLGGMOMLPAAAGADPLCRNPASRLKG
KLNSKILSNADCPKAEAFMSPTLPSTARAPKRSKRTGTGLSPKSPLEKRP
CCQPLLALPHDRASSTQGGEDNSGGKKPKTELGAPSPQPEGEQOPTRAKQ
PGQASTYSKRRALSRGRKTHASPCCKGRATRRQOQVLPDPAPEIRLKYISSC
KRLRADSTPAFSPFVVEKDAYTTICTVNSPGEPAHPKPSSSAASSTSSSSLL
EPAGASLFTTFGGSVLQORPSLPSLTHLGPVYSKALSTCLVCCLCQNPANFKDLG
DLCGPYVHECLPKKPKLEKLEGTLEASLPLERTLKECSASTTAAAPTAT
ITPTVALGRLSRPDPADPAKQGLRTSARGLSRRLOSCYCCDGGGGEVQAQDKS
RHEKSEAPTEFGGTQSEHWHEACAVTISGVYLVAGLGLQEAQMKVADWVRG"
7202..7207
polyA_signal
polyA_signal 7208..7213
polyA_site 7222
BASE COUNT 1612 a 2335 c 2027 g 1248 t
ORIGIN
alignment_scores:
Quality: 7504.50 Length: 1762
Ratio: 4.621 Gaps: 15
Percent Similarity: 92.168 Percent Identity: 81.725
alignment_block:
US-09-508-821B-6 x MUSMA ..
Align seg 1/1 to: MUSMA from: 1 to: 7222
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsnTy 17
309 ATGCAGTCTCTTTTCGAGAAAGGTGTGTTTCCATGCAACAGCAAACTA 358
17 xGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyArgGlnProS 34
359 CCCACAGACTCCCGAGGAGACATCGCGCTGGAGACTACAGCAGCCGG 408
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTy 50
409 CTCAGGCTGGCTAAGCTGTGATCGGCGAGCGGCTGCTGGCAAGACTAC 458
51 TyrAsnProGlnProTyrProSerTyrGluGlyAlaGlyThrProSe 67

966 rThrSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProG 983
3155 CACCTCACAGGCTCCTTGGGCCGGAAGCTTACAAAGCCTGCTGTACTCGT 3204
983 luAlaProIleAlaLysLysGluProValProAlaArgLysSerLeuArg 999
3205 AGGGGCCCAATTGTCTAAGAAAGAGCTGTGCCACGGGGTAAGAGCTTACGG 3254
1080. SerArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArg 1016
3216 rArgProIleHisAlaLeuLysArgLysSerAlaPheMetAlaProValP 1233
3225 AGCCGCGAGTACACCGGGGCTGCTTGGGCGGAAGACACTCTCCGTGCGAG 3304
1016 gAlaProValLeuProLysAspLeuLeuLeuProGluSerCysThrGlyP 1033
3305 GTGGCCAGCACTCCCAAAAGACTTGTGCTCCAGACTCTTGCACAGGCG 3354
1033 roProGlnGlyGluMetGluGlyAlaGlyValaProGlyArgGlyAlaSer 1049
3355 CCCCACAGACGACGCCGAGGGGCTGGAGCCCAAGCGCGGGGCTGTCA 3404
1050. GluGlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluPr 1066
3405. GAAGGCGCTCCCAAGATGTGACCGCTGCTTACAGCTCTGAGTAGGCC 3454
1066 oArgThrProGlyProProGlyLeuThrThrThrProAlaProProAspL 1083
3455 CCAGACTCTGGACCTTCGGGCGCTGACTACCAACCCCAAGCCTCCGACA 3504
1083 ySLeuGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGly 1099
3505 AACTGGAGGTAGACAGCAGCTCCTTCAAGCTTGGCAAGCGGGTGGGA 3554
1100 LysProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuPr 1116
3555 AAACCATCACCCAGAGCTCATCCAGCGCCACGACACCTGCTGCTCTGCC 3604
1116 oValAlaSerAspSerSerProMetGlySerLysThrLysGluThrAspS 1133
3605 TGTGTGCTTCAGACAGACGCCCCATCGGGCTCCAAAGACCAAGGCGCAGACT 3654
1133 erProSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArg 1149
3655 CTCCAGCATGCTGGCAAGGACCAAGCGCTCCATGTGCTCTCCGCTCTCGC 3704
1150 ThrLysThrGlnGluIlePheHisSerLysArgArgArgProSerGluG 1166
3705 ACCAAACCCACAGAGCTTCTCATGCCAAACGGCGGGCGGCTCCGAGAG 3754
1166 yArgLeuProAsnCysArgAlaThrLysLysLysLeuLeuAspAsnSerHisL 1183
3755 CCGATCCACAGACTGCGGTGCCACCAAGAGCTCCCTGCTTAAACACCAT 3804
1183 euProAlaThrPheLysValSerSerSerProGlnLysGluGlyArgVal 1199
3805 TACCCACTGGGTTCAGAGTCTCCAGTGGGCGCCACAGAGGAAGTCCGAGT 3854
1200 SerGlnArgAlaArgValProLysProGlyValaGlySerLysLeuSerAs 1216
3855 AGCCAGCGGGTCAAAAGTACCAAGCGTGGTACAGGAACAAGCTTTCAGA 3904
1216. pArgProIleHisAlaLeuLysArgLysSerAlaPheMetAlaProValP 1233
3905 TCGGCGCTTCACACGCTCCAGAGGAAGTGGCTTTCATGGGCACCTGTCC 3954
1233 roThrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSer 1249
3955 CCGCCAAAACGAGCCTCATCTCTGAGA. AGCAAC 3989
1250. AsnAlaSerAlaMetGly_GluMetGlyLysArgArgGlyLeuArgValP 1266
3990 AATGGAAAT. GGGGGAGACGGGAGGAGAGGGCTCAGAGATC 4033
1266 roProProSerSerArgGlyCysLeuLeuProAlaGlyProSerProPro 1282


```

4034 CCTGAGCTCTTACGAGAGATGCTCACCACCCAGAGGAGACGACGAGG 4083
1283 ArgAlaMetAlaSerLeuProGlnSerHisProArgArgProProPh 1299
4084 GTATG.....GGAGCCACCCCAACC..CCACCTT 4114
1299 eLeuProGlnAla.....ArgLeuSerAlaAlaPheGln 1311
4115 GGAGCCACTGCTGCATGCATGGGGTTGCTACAGAGCATCTCCCTCCA 4164
1311 LysAlaMetLysThrLysValLeuProProArgLysGlyArgLysLys 1327
4165 GTGCTGTGAGGACCAAGTCTGCTGCCACCCGAAAGGCGCCGCTCCAG 4214
1328 LeuGlnAlaIleValGlnLysIleThrSerProSerLeuLysLysPheAl 1344
4215 CTGGAGGCCATGATACAGAAATACCTCACTCGGCTCAAGAAATCTGC 4264
1344 acLysLysAlaProGlyAlaSerProGlyLysProLeuSerProSerLeu 1361
4265 ATGCAGAGGTGGAGGCGCCCTCTGAGGACACCCGAGACCCAGCCCTGC 4314
1361 eAspLysAspArgGlyLeuLysGlyAlaGlyLysSerProValGlyAl 1377
4315 CTGAG.....AGAGTCCAGGGGAGGAGCCAGCCGAGGAGG 4349
1378 GlnGlnGlyLeuValAsnValGlyThrGlyGlnLysLeuPro...ThrSe 1393
4350 GAAGAGGGTTTA.....GGAGGATGGGGCAAAATGCTACACAGCACTTC 4393
1393 rGlyAlaAspProLeuLysAspGlnProThrAspArgSerLeuLysGly 1410
4394 CGGAGCTGACCTATGTGCGAGAAACCAAGCAGAGCTGCTTAAAGTGA 4443
1410 LysLeuMetAsnSerLysLysLeuSerSer...ThrAspCysPheLysThr 1425
4444 AACTCTTGAACAGTAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4493
1426 GlnAlaPheThrSerProGlnAlaLeuGlnProGlyGlyThrAlaLeuAl 1442
4494 GAGGCTTCAATGTCCTCCGAGAGACCTGCTCATCTAGGAGCTCCCGAGC 4533
1442 aProLysLysArgSerArgLysGlyArgAlaGlyAlaHisGlyLeuSerL 1459
4544 ACCGAGAGAAAGAGAGCGGAAAGGACAGAGACTGTGGACCTCGACCTCA 4593
1459 LysGlyProLeuGlnLysArgProLysLeuGlyProAlaLeuLeuThr 1475
4594 AAGGTCGCTGGAGAGAGCGGCTGCGGACCACTGCTGCTGCTGCTGCT 4643
1476 ProArgAspArgAlaSerGlyThrGlnGlyAlaSerGlnAspAsnSerG 1492
4644 CCCACGACAGGGCCAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGG 4693
1492 GlnGlyGlyLysLysProLysMetGlnLysGlnLysGlnLysGlnLys 1509
4694 TGGAGAGGAGCAAGAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4743
1509 rOProGlnGlyArgProCysGlnProGlnThrArgAlaGlnLysGlnPro 1525
4744 CCCGAGAGGCGCGCTGCGACCCCAAGACCAAGGCGCGAGAGAGAGAG 4793
1526 GlyHisThrAsnLysSerSerLysSerLysArgLysArgLeuThrArg 1542
4794 GGCCAGAGCAGCTACAGAGCTATTCAGCAGGAGGAGGAGGAGGAGG 4843
1542 ValArgAlaLysAsnThrThrSerSerProCysLysGlyValArgAlaLysArg 1559
4844 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4893
1559 rGArgGlnGlnGlnValLeuProLeuAspProAlaGlnProGlnLysArg 1575

```

```

4894 GAAGGACGACGAGGAGTACTGCTCCCTGAGTCTGAGAGAGCTGAATCCGA 4943
1576 LeuLysThrIleSerSerCysLysArgLeuArgSerAspArgThrPro 1592
4944 CTCAAATACATTCCTCTCTGAGAGAGGCTGGGAGAGAGAGAGAGAGAG 4993
1592 AlaPheSerProPheValArgValGlnLysArgAspAlaPheThrThr 1609
4994 AGCTTCTGCGCTTGTGGGGGTGGAGAGAGAGAGAGAGAGAGAGAGAG 5043
1609 LeuThrValValAsnSerProGlyAspAlaProLysProHisArgLys 1625
5044 TATCACTGCTGCTCAACTCCCGGGGAGAGAGAGAGAGAGAGAGAGAG 5093
1626 ProSerSerAlaSerSerSerSerSerSerSerSerPheSerLeuAs 1642
5094 CCAATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5140
1642 PAlaAlaGlyAlaSerLeuAlaThrLeuProGlyLysSerIleLeuGln 1659
5141 ACCAGCTGGGGCTTCTGACACCAATTCCTGAGAGAGAGAGAGAGAG 5190
1659 rOArgProSerLeuProLeuSerSerThrMetHisLeuGlyProValVal 1675
5191 AGAGGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5240
1676 SerLysAlaLeuSerThrSerCysLeuValCysCysLeuCysGlnAsn 1692
5241 TCCAGAGCCCTTAAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 5290
1692 AlaAsnPheLysAspLeuGlnLysLysLysLysLysLysLysLysLys 1709
5291 GCCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5340
1709 LysLeuProLysLysLysLysLysLysLysLysLysLysLysLysLys 1725
5341 ACTGCTCCGCCCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5390
1726 GlyThrCysGlnGlnAlaSerLeuProLeuGlnLysThrLeuLysGly 1742
5391 GGCACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5440
1742 GlnCysAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1753
5441 GGAATGTTGGCCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5474
seq_name: gb-Pr.HSM801433
seq_documentation_block:
LOCUS HSM801433 2660 bp mRNA 18-FEB-2000
DEFINITION Homo sapiens mRNA, cDNA DKFZp434A139 (from clone DKFZp434A139).
ACCESSION AL133649
VERSION AL133649.1 GI:65992191
KEYWORDS
SOURCE
ORGANISM
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 2660)
AUTHORS Poustka, A., Klein, M., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) MIPs, Am Kioferspitze 18a, D-82152
Martinsried, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp434A139) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.

```


1266 AATGCTGCCAAGATACCTCTGCAAGCAGGTTGGCAG...TACAA 1312
322 sGlnProAspAlaValargThrProGluGlnTyrThrGlnThrPheS 339
1313 CCAGCTGAGGTCTCTAGGTCTCCAGTCTCCACCAAGATTCA 1362
339 erProSerSerHisSerProAlaArgSerValGlyArgSerProSer 355
1363 GCCTATTCTAACCTCTCCAGTCTCTGCTGTTGTTGCTCAGTCTCCAGC 1412
356 TyrSerSerThrProSerProLeuMetProAsnLeuGluAsnProTy 372
1413 TGAGCTCTACACCTCTCTCTCATGAGTGGTGGAGAACTCCAGTG 1462
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389
1463 TGGGCAA.....GGCAATGTCCTCAGTCTCCAGAAACC 1497
389 is...SerHisPheMetProLeuLeuAsnProSerProThrAspAlaThr 404
1498 GAATTTAGAGTATGCCCCAGCTCAGTCCCACTCCATCA...ATGATG 1544
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLysAspLy 421
1545 CCAGTCTTAACCTCTCATGCTGCAGGATTCAAAGGATTTGGATTAGAAG 1594
421 sLeuProGluLeuLeuSerAspLeuSerLeuGlnSerLeuThrAlaL 438
1595 GGTCCGGAAGAGCGGTGACTGATCTCTGGTGGTGGTGGTGGTGGTGGT 1644
438 eUrThrLeuGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeu 454
1645 TGAGTACTCAAGTGGCCATCTTCCATTAATCTGTCACACATGTTACTT 1694
455 SerLysAlaValProGlnLysLysGlyValLysAsnLeuValSerAr 471
1695 TGTGATGCTTGACACCTCAGAGAGAGGCTCTAAGAGGCGCTCATCA... 1742
471 gThrProGluGlnHisLysSerGlnHisCys...SerProGluGlySerG 487
1743TCATTAAGAAGCAGATAGCTGTACAACTCAGAGGCTCT 1785
487 LyTyrSerAlaGluProAlaGlyThrProLeuSerGluProProSerSer 503
1786 CACAGCTGAGACCAACTAAGTCCCTATG..... 1817
504 ThrProGlnSerThrHisAlaGluProGlnGluAlaAspTyrLeuSerG 520
1818GCTGAGTCTATTGGATGGAGTTCCTCTAGTAG 1849
520 ySerGluAspProLeuGluArgSerPheLeuTyrCysAsnGlnAlaArg 537
1850 TTCAGAGATCAAGGTGAGAG..... 1871
537 LySerProAlaArgValAsnSerAsnSerLysAlaLysProGluSerVal 553
1871 1871
554 SerThrCysSerValThrSerProAspAspMetSerThrLysSerAspAs 570
1872GTGAGGCAACTGAGTGGCCAGAGCAGTCTGTGACAC 1909
570 pSerPheGlnSerLeuHisGlySerLeuProLeuAspSerPheSerLysp 587
1910 CACCTATAGTGT.....GGCGCTCAGAGAAAGTGGCTCTCCACAA 1953
587 heValAlaGlyLysArgAspCysProArgLeuLeuLeuSerAlaLeuAla 603
1954 CACAAGGTGCTCAGAAAGAGGCGCTTACCTCAGTACCTGCTGCACT 2003
604 GlnGluAspLeuAlaSerGlu..... 610
2004 AGGAGCAGAGTGCCTCTCCAGGTGCTAAGGACACATCATCTCATCTGA 2053

641IleLeuGlyLeuGlnG 616
2034 GGGGAACACAAAATCAATGAGAAGACAGTTGGGTGATTGCTCCGAG 2103
646 luAlaIleGlyGluLysAlaAspLysAlaTyrAlaGluAlaProSerLeu 632
2104, AAGCATGACAGTGGGTGAGAAATCTGTGGACAGATAAGGCTCC 2153
633, VallysAsp.....SerSerLysProPheSerLeuGluAs 645
2154, CAGAGATGATCTCTGTCAGTCCAGAGACCACT..... 2189
645 rHisSerAlaCysLeuAspSerValAlaLysSerAlaTyrProArgProG 662
2190, AGCAATAGTGGGTAAAGAAATCAGCCACACATCTCTCCAGCCAG 2238
662, LyGluProGluAlaLeuProAspSerLeuGlnLeuAspLysGlyAsn 678
2239, AFCTCTCAGGAGA.....GGGAGCAAGGAACAAGATGGTATAT 2282
678, AlAlaLysAspPheSerProGlyLeuPheGluAspProSerValAlaPheAl 695
2283, AATAGCTTAACCAATGAGAGGGAATGGCCGAGTAGCCACTCCGC 2332
695, aThr.....ProAspProLysLysThrThrGlyP 705
2333, AGTAGGCCAAGTTTCACAGGAGGATGAGCTCAGAGTCTCTCGAA 2382
705, rLeuSerPheGlyThrLysProThrLeuGlyValProAlaProAspPro 721
2383, GTTCCGCTATGATTAAGAGAGTTTGGTCACTCTACCA...CGA 2429
722 ThrThrAlaAlaPheAspCysPheProAspThrThrAlaAlaSerSerAl 738
2430, AATGTCAGTGTATCTCATGATCTCT.....TCAGG 2461
738, aspSerAlaAsnProPheAlaTyrProGluGluAsnLeuGly..... 752
2462, ACAAGAAAGGGGATTTGGCAGCCATGGAGAGCGAAGAGGTAGAAATG 2511
753AspAlaCysProArgTyrGlyLeuHis 761
2512, AGAGTCTCCAGTCTTCTACAGGAGTGTCTCAGGGTACCCACCACAT 2561
762, ProGlyGlu.....LeuThrLy 767
2562, CCTGACAGAGGATCTCTAGAGCCGACAGGAACATCAAGGGATGGCTAG 2611
767, sGlyLeuGlnGlnGlyLysAlaSerAspGlyLysSerLysGlyAspT 784
2612, TGGCTGGAA.....GGAATGCAAGGCTTAACATCTTAGTCACTCAA 2655
784, hrHisGluAlaSerAlaCysLeuGlyPheGlnGluGluAspProGly 800
2656, CCAATGAATTAGTAGC..... 2672
801, GluLysValAlaSerLeuProGlyAspPheLysGlnGluValGlyG 817
2673AGGGGCT 2680
817, yValLysGluGluAlaGlyLysLeuLeuGlnCysProGluValAlaLysA 834
2681, TCTGAACAAGAGTATGGTATCCCTGTGGAAATCCC..... 2717
834, laAspArgTyrLeuGluAspSerArgHisCysCysSerThrAla..... 848
2718, CACTGGGACATGGGAAGGAAGTCAAGCAGCAGCTCTGNA 2762
849AspPheGlyAspLeuProLeuLeuProThrSerAr 861
2763, ATGAACAGATCAATTTCTGTGATCTCATCATT.....CCGAG 2800


```
4424 ACTATCTCTAAAGAGTGTCTCTCTGAGGTGGACACTGTGGCTACTCAAG 4473
1377 .....ValGluGluGlyLeuValAsn 1383
4474 AGGTGAGATGGAAGAAAGAAAGTGTGAGGTGTATCTGACCTGTGCTAGT 4523
1384 ValGlyThrGlyGlnLeuProThrSerGlyAlaAspProLeuCysAr 1400
4524 GTA...ACTACCAAGAA...TCGAATGTGAGAAGCCTCTTCC 4561
1400 GasnProThrAsnArgSerLeuGlyGlyLeuMetAsnSerLysLysL 1417
4562 AGGCGCT...TCTGAAGAGTGGCGTGC... 4586
1417 euSerSerThrAspCysPheLysThrGluAlaPhe...ThrSerPro 1431
4587 ..AGTGGGGATGACAAAGTCAAGACAGAGGCACATGTAGAACAGCTTCT 4634
1432 GluAlaLeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerAr 1448
4635 ACTGGAAGGAAGAACCTCTGTGTACTATACATCCACAGCTTCACAGAAGCC 4684
1448 glyGlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysA 1465
4685 TGGTGTAAACCAAGGAGACAGATGCTTCCCTGGGT... 4721
1465 rgProTyLeuGlyProAlaLeuLeuLeuThrProArgAspArgAlaSer 1481
4722 .....GGGCGACACACACTATATCTTCTTCCCTCAAGAAGATGTA 4760
1482 GlyThrGlnGly...AlaSerGluAspAsnSerGlyGlyGlyGly 1496
4761 GTCACAGTGGCATATGCGCCCTGAGCAATCCCAAGGCTGAAGAGAA 4810
1496 sLys...ProLysMetGluGluLeuGlyProA 1506
4811 AGAAGACGATACACTCATGATTTCCCAACACAGAAAGTGTCCGCCCTA 4860
1506 laSerGlnProPro...GluGlyArgPro...Cys 1515
4861 AAGGGTATTCCCTACGAGAAAGAAAGGGGAGCAACATCGTAGTGTG 4910
1516 GluProGlnThrArgAlaGlnLysGlnPro... 1525
4911 AATAAGCAAGAAACACAGCAGCAGCACCCTCCCTCCCTCAGCCGCC 4960
1526 .....GlyHisThrAsnTySerSerTySerLysArgLysA 1538
4961 TCAGTCCGACAGAGGTTCTCAGATGGGAGCCAAAGCCAAAGACAGA 5010
1538 rgLeuThrArgGlyArgAlaLysAsnThrSerSerProCysLysGly 1554
5011 GGCAAGAGGAGGAGAGAGGAG... 5045
1555 ArgAlaLysArgArgArgGlnGlnGlnValLeuProLeuAspProAla 1571
5046 CCAAGAGAGGAGAAACCAACACAGCAGTCCCTTGTAGAACCTCAAGA 5095
1571 uProGluLeuArgLeuLysTyLysSerSerCysLysArgLeuArgSerA 1588
5096 ACCAGATCAAGTCAAGTAAAGTACGTACCCAG...CCACTAGATAAAACTG 5142
1588 sSerArgThrProAlaPheSerProPheValArgValGluLysArgGsp 1604
5143 ATGCCAAGACAAAGTCTTTTCCCTTACATCCATGATAGTAATAGTGT 5192
1605 AlaPheThrThrileCysThrValValAsnSerProGlyAspAlaProLy 1621
5193 GAAGTGGAGCTGTGTGTACATCATCAATGCTGAAGAGAGAACAGAC 5242
1621 sPro...HisArgLysProSerSerSerSerSerSerSerS 1635
5243 CAAATTTGTTGAGGAGCGGAGGAGGTCAGAGATCTCTGACCCCTCC 5292
```

```
1635 erSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAlaThrLeu 1651
5293..GCAGCAGAAAGCAAGGTG..... 5312
1652..ProGlyGlySerIleLeuGlnProArgProSerLeuProLeuSerSerTh 1668
5313.....CTCCAGCTTCATCCTT 5329
1668..rMethHisLeuGlyProValValSerLysAlaLeuSerThrSerCysLeuV 1685
5330..TATGCTGTCAGGGGCTGTGTAAACAGAGTCTTCTGTATTGGGGCAGCTGG 5379
1685..aLysCysLeuGlyGlnAsnProAlaAsnPheLysAspLeuGlyAspLeu 1701
5380..TTTGTCTGTCTGTGGCAAGTGGCCAGTTACCGTAAACATGAGGTGACCTC 5429
1702..CysGlyProTyTyTyProGluHis.....CysLeuProLysLysLys 1715
5430..TTTGGACCTTTTATCCCAAGATTATCCAGCCACTCTTCCGAGAATCC 5479
1715..sPro.....LysLeuLysGluLysValArgProG 1725
5480..ACCTCTAAGAGGTCTCTAGAAATGCAGACAAAGTCAAGGTTCGGCAC. 5528
1725..luGlyThrCysGluGluAlaSerLeuProLeuGluArgThrLeuLys.G 1741
5529.....AAAGCGCTCTTATGTTCTTAAACTGCACACTGAGGAG 5567
1741..LyProGluCysAlaAlaAlaThrAlaGlyLysPro 1753
5568..GAGGAGGAGCAGCAGCAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5605
```

seq_name: gb_pr:HS597B2

```
seq_documentation_block:
LOCUS HS597B2 43761 bp DNA PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP4-597B2 on chromosome 22 Contains
an exon of the TCF20 gene for transcription factor 20
(ARI,KIAA0292), ESTs and GSSs, complete sequence.
ACCESSION AL031346
VERSION AL031346.8 GI:4760290
KEYWORDS HTG; ARI; KIAA0292; TCF20.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 43761)
Babbage, A.
Direct Submission
Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 7, 1999 this sequence version replaced gi:4582316.
This sequence has been finished according to sequencing map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
```

FEATURES
source

```

repeat_region 1121..1159 /note="13 copies 3 mer ctc 7c conserved"
misc_feature complement(2687..3321)
/gene="TCRF20"
/note="match: GSS: Em:AQ606694"
5484..5525 /note="14 copies 3 mer tgc 7g conserved"
6508..6516 /note="14 copies 3 mer tgc 7g conserved"
7232..7400 /note="Alu repeat: matches 1..306 of consensus"
/note="FAM repeat: matches 8..176 of consensus"
8240..8712 /note="L2 repeat: matches 2138..2688 of consensus"
9039..9433 /note="match: GSS: Em:AQ119461"
9222..9342 /note="MIR repeat: matches 113..252 of consensus"
10256..10539 /note="Alu repeat: matches 1..312 of consensus"
complement(10280..10631) /note="match: GSS: Em:AQ554699"
10540..10563 /note="12 copies 2 mer aa 100 conserved"
10757..11055 /note="Alu repeat: matches 1..299 of consensus"
11358..11672 /note="Alu repeat: matches 1..294 of consensus"
11780..12189 /note="LTR33 repeat: matches 7..466 of consensus"
12575..12641 /note="LTR33 repeat: matches 454..519 of consensus"
12699..12835 /note="MER6 repeat: matches 1..160 of consensus"
12923..13230 /note="Alu repeat: matches 1..308 of consensus"
14064..14356 /note="Alu repeat: matches 1..296 of consensus"
14637..14784 /note="MIR repeat: matches 100..262 of consensus"
15383..15668 /note="Alu repeat: matches 1..291 of consensus"
15726..15823 /note="49 copies 2 mer aa 65 conserved"
15884..15931 /note="L1ME repeat: matches 5759..5805 of consensus"
16192..16486 /note="Alu repeat: matches 1..301 of consensus"
16525..16834 /note="Alu repeat: matches 1..310 of consensus"
complement(16676..17109) /note="match: GSS: Em:AQ698298"
17613..17930 /note="Alu repeat: matches 1..310 of consensus"
18031..18170 /note="L1M4 repeat: matches 4457..4612 of consensus"
18193..18477 /note="Alu repeat: matches 1..291 of consensus"
18493..18799 /note="L1M4 repeat: matches 7630..7972 of consensus"
18832..18938 /note="L1M4 repeat: matches 7445..7560 of consensus"
19080..19145 /note="MER5A repeat: matches 44..112 of consensus"
19226..19516 /note="L2 repeat: matches 1921..2236 of consensus"
19543..19716 /note="MIR repeat: matches 79..256 of consensus"
19755..20002 /note="L2 repeat: matches 1618..1833 of consensus"
20253..20387 /note="L2 repeat: matches 1456..1583 of consensus"
20633..20683 /note="L2 repeat: matches 2677..2729 of consensus"

```



```

repeat_region 21378..22070
/notes="finger3b repeat: matches 2..741 of consensus"
22071..22382
/notes="Alusx repeat: matches 1..309 of consensus"
22383..22393
/notes="finger3b repeat: matches 741..751 of consensus"
22394..22673
/notes="Alusx repeat: matches 1..282 of consensus"
22674..22840
/notes="finger3b repeat: matches 751..928 of consensus"
22841..23138
/notes="Alu repeat: matches 1..297 of consensus"
23139..23585
/notes="finger3b repeat: matches 928..1231 of consensus"
23738..24015
/notes="Alu repeat: matches 1..313 of consensus"
24067..24116

```

alignment_scores:

```

Quality: 993.50 Length: 2067
Ratio: 1.027 Gaps: 96
Percent Similarity: 46.783 Percent Identity: 25.109

```

alignment_block:

US-09-508-821B-6 x HS597B2/rev ..

Align seg 1/1 to reverse of: HS597B2 from: 1 to: 43761

```

1 MetGlnSerPheArgGluGlyCysGlyPheHisGlyLysGlnGlnAsnTy 17
|||||
6451 ATGCAGTCCCTTCGGGACAAAGCAGTTACCCAGGAACACGAAAGCTA 6402

17 rGlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyArgGlnPro 34
|||||
6401 CCCACAGAGGAGTACAGGCTCATCCGCTTGAAGAGTTC...AGCCCTC 6355

34 exGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTyr 50
|||||
6354 GTCAGGCC..... 6347

51 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyThrProSe 67
|||||
6346 .....CAGATGTTCCAGATTTT...GGAGGTACAGGTGCCAGTAG 6309

67 rGlyThrAlaAlaAlaAlaAlaAlaAspLysTyrHisArgGlySerLys 84
|||||
6308 TGCCAGCAGTGGCAGT...GGCAGTGGTGGTGACACGAGCAGCAGCAG 6262

84 lAlaLeuProThrGlnGlnGlyLeuGln.....GlyArgProAlaPhePro 98
|||||
6261 CTGCTGCCGAGCGAGTGGCTACGAGACCTCTGCCATCAAGTTACCAG 6212

99 GlyTyrGlyValGlnAspSerSerProTyrProGlyArgTyrAlaGly 115
|||||
6211 GGTTCAGGAAGAGGCTGGAGATTTTAC.....TACATGGCAGGCAA 6168

115 uGluSerLeuGlnAlaTrpGlyAlaProGlnProProProGlnProG 132
|||||
6167 CAAGAAGCCCGTACTACAGAACCCACAGCCT.....CCTC 6130

132 lProLeuProAlaGly...ValAlaLysTyrAspGlnAsnLeuMetLys 147
|||||
6129 AGGAGAGGCTCTTGGCCCTGTGTCAGAGCTATGA..... 6095

148 LysThrAlaValProPro.....SerArgLntTyrAlaGly 159
|||||
6094 .....CCCCCCCAGGGAGCAGCTTTGGCAATCAGTATGGAG 6057

159 uGlnGly.....AlaGln..... 163
|||||
6056 TGAGGTGTCATGTGGCCAGGTTTCAACGACAGCAGCTCTGGCCTTGGCGTG 6007

164 .....ValProPheArgThrHisSerLeu 171

```

```

6006 TGTCACATTATCAGCAGGATTACACTGGGCGCTTCTCTCCAGGAGTGCT 5957
|||||
172 HisValGlnGlnProProProGlnGlnProLeuAlaLysTyrProLysLe 188
|||||
5956 CAGTACCAACAGCAGGCTTCAGCCAGCAGCAGCAGCAGCAGCAGCAGCA 5907
|||||
188 uGlnArgGlnLysLeuGlnAsnAspLeuAlaSerProLeuProPhePro 205
|||||
5906 GTTGAGAGCAACAGCTTTACCAG.....TCCCATCAGCCCTGCCAC 5866
|||||
295 lGlyThrHisPhePro.....GlnHisSerGlnSerPhePro 217
|||||
5865 AGGCCATGTGCCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5816
|||||
218 ThrSerSerThrTyrSerSerValGlnGly.....Gln 229
|||||
5815 CGGCCCTCACTCTGCCATCCTCTGCTGCTGTACCAAGTTAAGAGTGGG 5766
|||||
229 yGlyGlnGlyAlaHisSerTyrLysSerCysThrAlaProThrAlaGlnP 246
|||||
5765 TCAGTTTGGCCAAACACTATCAGTCTTCTCTCTCTCTCTCTCTCTCT 5716
|||||
246 rHisAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGln... 261
|||||
5715 CCTTCCCTTACCACAGCGTTTTAGCCAGTCT.....GGCAGAGC 5675
|||||
261 ..... 261

5674 TATGATGCCAGTTACATATGTAATCTGGATCTCAGTATGAAGGACACAA 5625
|||||
262 .....A 262

5624 TGTGGTCTTAATGCACAGGCTTATGGAACACATCCAATTACAGCTATC 5575
|||||
262 rGlyValGlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGln 278
|||||
5574 AGCCTCACTATCAAGATTTTGACAGGCAAGATTCACACAGGGACC 5525
|||||
278 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 295
|||||
5524 CAACAGGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 5486
|||||
295 rArgHisHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAla...LysT 311
|||||
5485 .....CACCCCTCTCAGCATGTGTGCGATATACAGCTGCCACCCAGC 5440
|||||
311 yGlnHisTyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 327
|||||
5439 TGCCCTCGAAGCCAGCTGGGCGAG...TACACAGCAGCTGAGGTCTCT 5393
|||||
328 ValArgThrProGlnGlnTyrTyrGlnThrPheSerProSerSerSerH 344
|||||
5392 GTGAGGTGCCCATCGATGCTTCCACAGAACTCAGCCCATTTCTAACC 5343
|||||
344 sSerProAlaArgSerValGlyArgSerSerTyrSerSerTyrProS 361
|||||
5342 TTCTCAGCTGCTCTGTGCTGCTCCAAAGCTGATGTTCTACCCCAT 5293
|||||
361 rProLeuMetProAsnLeuGlnAsnPheProTyrSerGlnGlnProLeu 377
|||||
5292 CTCCTCTCATGAGCAGCTGGGAGATCTCCAGTGTGGGCAA..... 5252
|||||
378 SerThrGlyAlaPheProAlaGlyLysThrAspHis...SerHisPhe 393
|||||
5251 .....GGCAGTGTCCATGTTCCAGAAACAGATTTTACAGTTAAT 5208
|||||
393 rProLeuAsnProSerProThrAspAlaThrSerSerValAspThrG 410
|||||
5207 GCTCAACTCAGTCCACCCATCA...ATGATGCCAGTCCCTAATCTC 5161
|||||
410 lAlaGlyAsnCysLysProLeuGlnLysAspLysLeuProGluAsnLeu 426
|||||

```


5160 ATGCTGAGGCTTCAGAGGCTTGACACAGAGGCTACCAAGGAGGCA 5111
 427 LeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuThrLeuGlnValG1 443
 5110 CTGACAGATCTCGGTGAGAGTTGAGTGTCTGCTGAGTACCAAGTGGC 5061
 443 uAsnIleSerAsnThrValGlnGlnLeuLeuSerLysAlaAlaValP 460
 5060 CAATCTCTTACACTGTCAGCAGCATGTTACTTCTGAGTCCGCAATC 5011
 460 roGlnLysLysGlyValLysAsnLeuValSerArgThrProGlnGlnHis 476
 5010 CTCAGAGAGAGACCTCCAG.....AGGCCCTCATCTTCCAG 4973
 477 LysSerGlnHisCys...SerProGlnLysSerGlyLysSerAlaGluP 492
 4972 AAACAGAGATAGCTGCACAAATCTGAGAGGCTCTCAGCACTGAGAAC 4923
 492 oAlaGlyThrProLeuSerGlu.....ProProSerS 503
 4922 GCTGAGTCCCTATGCGAGATCATTAATGAGAGGCTGCTCCAGCAATT 4873
 503 erThrProGln.....SerThrAlaGluPro 512
 4872 CAGAGATCAAGGCGAGAGATGCGCAACTAGTGGCCAGAGCAACGAC 4823
 513 GlnGlnAlaAspTyrLeuSerGlySerGlnsProLeuGlnArgSerPh 529
 4822 TCTGACACCACTCAAGAGGAGGCTTGAGAAAGCTGCTCTCTCA.. 4775
 529 eleuTyrCysAsnGlnAlaArgLysSerProAlaArgValAsnSerAs 546
 4774CCGGCCAAAGGTGCTCGAATGAACCCCGACCTCAATGCTATGC 4729
 546 erLysAlaLysProGlnSerValSerThyCysSerValThrSerProsp 562
 4728 CTGCCGCAAGAGAGAGGCC.....ACCTGACCA... 4700
 563 AspMetSerThrLysSerAspSerPheGlnSerLeuHisGlySerLe 579
 4700 4700
 579 uProLeuAspSerPheSerLysPheValAlaGlyLysArgSerProA 596
 4699GGCGTAAAGGATGCCA. 4682
 596 rgleuLeuSerSerAlaLeuAlaGlnLysLeuAlaSerGlnLeu 612
 4681TTGTCAATCCGAGGAGAACCCAAAGGTTATGAGAAAGCTTT 4640
 613 GlyLeu.....GlnGlnAlaIleGlyLysAlaLysAlaTr 626
 4639 GGGGTGATGTCTCCGGGAGGACAGACAGTGGGTAAAGAACCTGG 4590
 626 PALGlnAlaProSerLeuValLysAsp.....SerSerLysP 639
 4589 TGGACAGATTAAGGCTCCCAAGAGATATCTCGACGACCTCAAGGC 4540
 639 roPheSerLeuGlnAsnHisSerAlaCysLeuAspSerValAlaLys 655
 4539 CACCT.....AGCAATGCTGGGGCAAAGAAACCACTATCATCA 4499
 656 SerAlaTrpProArgPro.....GlyGluProG1 665
 4498 CTTCGCCAGCCAGACCTCCAGAGAGAGAGGAGCAAGAAAGAAAGAA 4449
 665 uAlaLeuProAspSerLeuGlnLeuAspLysGlyLysAlaLysAsp 682
 4448 TGGCGATTAACAACCTCAACATATGAGAAAGAAAGGCGAGATGGCC 4399
 682 heser.....ProGlyLeuPheGluAspProSerValAlaPheAla 695
 4398 ACTCTGCAAGGGGCGCCCTGTTT.....ACGAGC 4370

696 ThrProAspProLysLysThrThrGlyProLeuSerPheGlyThrLysP 712
 4369 AGACAGACCTACCAAAATCTCTGAGAGCTGCGCTATGATACAAAG 4320
 712 cThrLeuGlyValProAlaProAspProThrThrAlaAlaPheAspCysP 729
 4319 TAGTTGGGGTCAACCGCTGCA...CGAAATGCAATGGCTTCTTCAGT 4273
 729 heProAspThrThrAlaAlaSerSerAlaAspSerAlaAspProPheAla 745
 4272 ATCTT.....ACAGGGCAAGAAAGGAGAT.....TTCACT 4241
 746 TrpProGlnGlnAsnLeuGlyAspAlaCysProArgTyrPheLysHisP 762
 4240 GGCCATGGGGAAACGAAAGGAT.....AGAAATGAAATATCC 4203
 762 oGlyGlnLeuThrLysGlyLeuGlnGlnGlyLysAlaSerAspLysP 779
 4202 AAGCTCTCGAGAGAGTCTTCAGGTTACACCAACCTACAGAG 4153
 779 leSerLysGlyAspThrHisGlnAlaSerAlaCysLeuGlyPheGlnGlu 795
 4152 GATATTCTGAGTACTCAAGACATCAGGGAGTGGCTGTAAGCTAGAA 4103
 796 GluAspProProGlyLysLysValAlaSerLeuProGlyAspPheLysG1 812
 4102 GGAACCAAGGCCCAATGCTTGTTGTGTAACCAATGATTAAGTCTAG 4053
 812 nGlnGlnValGlyLysGlnGlnAlaGlnGlyLysLeuLeuGlnCysP 829
 4052 CAGG.....GGCCTCTGAACAAAGATGGGTGCTCATTAAGAAATG 4009
 829 roGluValAlaLysAlaAspArgTrpLeuGlnsPheArgHisCysG 845
 4008 CC.....CACTGGGGCCCTGGGAGAGGAAATCAAGC 3977
 846 SerThrAla.....AspPheGlyAspLeuProLeu 856
 3976 AGCAGACTCTGTAATGAATGAAAGATCAATATTGACTATACCAAT.. 3929
 856 uProProThrSerArgLysGlnAspLeuGlnAlaGlnGlnGlnLysP 873
 3928CCGAGAAAGTTGTAATAGACCTGAG..... 3902
 873 erLeuGlnLeuLeuGlySerProGlnGlnArgProGlyLysGlnAsp 889
 3901TCAACAGACATAGACCTGG.....GGT 3878
 890 ProLeuSerProLysAlaProLeuLecysThrLysGlnGlnValGlnG1 906
 3877 TCCTCTCTGAAAGATCATGATCTGTGATATTCTTCCACTAAGACA 3828
 906 uAlaLeuAspSerLysAlaGlyTrpLysProCysHisLeuSer.... 921
 3827 GATTGTC..AGGACCCAGGGGCTCATCTCTGAGACACATAGATGGC 3781
 922GlyGlnSer 924
 3780 ACAACCAATTTGGAGAGATGACCTCATCACTCACTTAAGTCAAGTCA 3731
 925 ValIleLeuLeuGlyProThrValAlaGlyHisGlnSerLysVal.... 938
 3730 GTCAATCTCTCTGTTGTTGTTGTCTCATGAGAAACCACTCAATATCCA 3681
 939GlnSerTrpPheGlnSerS 945
 3680 GAGCGGCGAGATTAAGAGAGAGACTTTGAGACAGTCAATATCAACTA 3631
 945 erLeuSerHisMetLysProGlyGlu..... 953
 3630 GTTTCACACACAGAAATCTGAGAGACCTGCACTCTCTTAAGATCAAG 3581

2066 TGCCATGACATCCACAACTCAGAGAGCCGTGTAATACCAAGGAGAC 2017
 1452 la...glyAlaHisGlyLeuSerTyrGlyProLeuGlnTyrSerProTyr 1467
 2016 CAGATGCTCCCTGGT 2000
 1468 LeuGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGlyThrG1 1484
 1999 ...GGAGACGACCTTAATCTTCACAGCCAAAGATATGATACCTCCAGT 1953
 1484 ngl.....AlaSerGlnAspAsnSerGlyGlyGlyGlyTyr... 1497
 1952 GGGCATATTGGCCCTGAGCAACCCAGAGCTGAGAGAGAGAGAGACG 1903
 1498ProLysMetGlnGlnLeuGlyProAlaSerGln 1508
 1902 ATACAGTACGATTTCCACAGAGAGAGGTTTCCCTCCAGAGGATAT 1853
 1509 ProPro.....GlnGlyArgProGlyGlnProGlnThrArgAl 1521
 1852 TTCCCATACAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1803
 1521 aglnTyrGln..... 1524
 1802 AAAGAAACAGACAGACAGCCAGCCCTCCAGCCCTCCAGCCCTCCAGATAC 1753
 1525 ..ProGlyHisThrAsnTyrSerSerTyrSerTyrSerTyrSerTyr 1540
 1752 CAGAGAGTTCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1703
 1541 ArgGlyArgAlaLeuAsnThrThrSerSerProCysGlyGlyArgAla 1557
 1702 AGGAGAGAGAGAGAG.....CTGGGGCCAGCCAGAGAGAGAGAG 1668
 1557 SAAGATGATGlnGlnGlnValLeuProLeuAspProAlaGlnProGln 1574
 1667 GCGAAAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1618
 1574 LeuArgLeuTyrTyrLeuSerSerCysLysArgLeuArgSerSerArg 1590
 1617 TCMACTAAATATCCACCCAG...CCACTGGATTAATATCTATCCAG 1571
 1591 ThrProAlaPheSerProPheValArgValGlnTyrArgAspAlaPhe 1607
 1570 AACAGCTTTTACCTTACATCATCATCATTAATATATGAGAGACTGG 1521
 1607 rThiLeuSerThrValValAsnSerProGlyAsp..... 1618
 1520 AGCCGTTGTACATCATCATCATCATCATCATCATCATCATCATCATCAT 1471
 1619AlaProLysProHisArgLysProSerSerSerAlaSerSer 1633
 1470 TGAAGGAG 1421
 1634 SerSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuAla 1550
 1420 GAAAGCAG..... 1412
 1650 rLeuProGlyGlySerIleLeuGlnProArgProSerLeuProLeuSer 1667
 1411GCGCTCCGCGCTCGT 1396
 1667 erThrMetHisLeuGlyProValValSerLysAlaLeuSerThrSerCys 1683
 1395 CTTTATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1346
 1684 LeuValCysCysLeuGlnAsnProAlaAsnPhelAspLeuGlnVal 1700
 1345 CTGGTTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1296
 1700 PLeuGlyGlyProTyrTyrProGlnHis.....CysLeuProLysL 1714
 1295 CTTCTTTGAG 1246

1714 yslYsPro.....LysLeuGlyGlyValArg 1723
 1245 ATCCACCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1196
 1724 ProGlnGlyThrCysGlnGlnIleSerLeuProLeuGlnArgThrLeu 1740
 1195 CAC.....AAAGTCTCTTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1158
 1740 ys.glyProGlnCysAlaAlaAlaAlaThrAlaGlyLysProPro 1754
 1157 GGAG 1113
 seq_name: gp_sts:AF021115

seq_documentation_block: 595 bp DNA STS 03-APR-1998
 LOCUS AF021115
 DEFINITION Homo sapiens trinucleotide repeat ctg-23, sequence tagged site.
 ACCESSION AF021115
 VERSION AF021115.1 GI:3015588
 KEYWORDS STS.
 SOURCES human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 595)
 Philbert,R.A., Horelli-Kutunen,N., Robb,A.S., Lee,Y.H.,
 Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,
 Palotie,A. and Gims,E.I.
 The characterization and sequence analysis of thirty CTG-repeat
 containing genomic cosmid clones
 Eur. J. Hum. Genet. 6 (1), 89-94 (1998)
 98454304
 2 (bases 1 to 595)
 Philbert,R.A., Horelli-Kutunen,N., Robb,A.S., Lee,Y.H.,
 Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,
 Palotie,A. and Gims,E.I.
 Direct Submission
 Submitted (28-AUG-1997) Clinical Neuroscience Branch, National
 Institutes of Mental Health, 49 Convent Drive Rm 31E16, Bethesda,
 MD 20892, USA

FEATURES
 source 1..595
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"
 /map="17q11.2"
 /note="primers to amplify repeat: forward:
 ACAGCAGAGTCCAGCCCAT; reverse:
 GGTGATCTGGGAGAGTTGGT"
 BASE COUNT 90 a 140 c 230 g 120 t 15 others
 ORIGIN

alignment_scores:
 quality: 873.50 length: 199
 ratio: 4.671 gaps: 6
 Percent Similarity: 93.970 Percent Identity: 88.945

alignment_block:
 US-09-508-821b-6 x AF021115/rev ..

Align seq 1/1 to reverse of: AF021115 from: 1 to: 595
 178 ProProGlnGlnProLeuAlaTyrProLysLeuGlnArgGlnLysLeu 194
 533 CCGGCCAG 544
 194 LAsnAspIleAlaSerProLeuProPheProGlnGlnLysHisPhePro 210
 543 AG 494
 211 GlnHisSerGlnSerPheProThrSerThrThrThrSerSerValG1 227

860 SerArgLysGluAspLeuGluAlaGluGluLysrSerSerLeuCysG1 876
1793 CCCAGAGATTGAAATAGAGCTCAG 1819
876 uLeuLeuGlySerProGluGlnArgProGlyMetGlnAspProLeuSerP 893
1820TCATCAGCATAGAGCTGGG GGTCCCTCTCTG 1853
893 roLysAlaProLeuLleCysThrLysGluGluValGluGluValLeuAsp 909
1854 AAGAAGATCAGTGAATCTGATGATTTCTCCACTAAGACAGATGTC ... 1900
910 SerLysAlaGlyTrpGlySerProCysHisLeuSer 921
1901 AGGGACCCAGGGGCTCACTCACTGGGACACATGATGCTCCGACACCAAGAT 1950
922GlyLysrValLleLeuL 928
1951 TGGGAGGAATGACCTTCATCAATCACTTAACTGATGCTGGTCTCTTC 2000
928 euGlyProThrValGlyThrGluSerLysVal 938
2001 CTGGTGGTTGGTGTCCATGGAACCACTGAATCCAGAGCGGGCAG 2050
939GlnSerTrpPheGluSerSerLeuSerHi 948
2051 AATAAGAGGAGACTTTGAACTGTAATCTCAAGTAGTTTCAACAA 2100
948 sMetLysProGlyGlu GluGlyP 956
2101 CAAGAATCTGGAGACCACTGCCATCTCTAGCTCAAGCATGAGTCT 2150
956 roAspGlyLysAlaProGlyAspSerThrThrSerAspAlaSer ... 971
2151 ACCGGGGAATGCGACCTCTGGGACGACCAACCATGATCTCTTCAGAC 2200
972 LeuAlaGlnLysProAsnLysProAla ValProGluAl 984
2201 TATGGCCGCGCAAGACAGCAGACCCACGCCAATCGCGGGTCCCT...GG 2247
984 aProLleAlaLysGluProValProArgGlyLysSerLeu 998
2248 CAGAGTTGGTGGTGGGAGGGCATG...AGGGGTGGTCCCTCTCAAT 2294
999ArgSerArg... 1001
2295 ATCACTGACTTTCAGAAAAATTGAAATGCTCTCTGGCGGAGCAGAGGC 2344
1002ArgValHisArgGlyLeuProGlu... 1009
2345 CCAGGGGAGACCTCATCATGAATCCACATGACCTTTTCAGAGAG 2394
1010 AlaGluAspSerProCysArgAlaProValLeuProLysAspLeuLeu 1026
2395 GGCTAACCGGAGTCTTTACACACTCCCTTTTCTCCCACTCAGAAACC 2444
1026 eu 1026
2445 TGGCTCTCTATCATGCAATCTCGGGCTCATGTTATGGGAGCCCT 2494
1026 1026
2495 AACGCAGTTGAAATCTCAGCTGCATTATAAGACAGATGTAACAACA 2544
1027ProGluSerCysThrGlyProGlnGlyMetGlnGlyAlaG 1042
2545 GCACACAGAGAGTAAAGACTGGAGCAGCGGTCTGCTCAGGGAGTAA 2594
1042 lValAlaProGlyArgGlyAlaSerGluGlyLeuProArgMetCysThrArg 1058
2595 TTGCTGCAGCAGCAGCAGGAGGAGGG...CCAGGAAGAGTCCAAAGG 2641

1059 SerLeuThrAlaLeuSerGluProArgThrPro 1069
2642 CAGCAGCAGTTCTTCAGACAGATGAGGAGCCCTCTGAAAAATGACAAAGA 2691
1070GlyProPro...GlyLeuThrThrProAlaPro 1080
2692 TGGTATGATGATGAGCCACCACTGGGACTTACCATGACCCCAAGTGGCC 2741
1081 ProAspLysLeuGlyGlyLysGlnArgAlaAlaPheLysSerGlyLysr 1097
2742AGGAG 2746
1097 gValGlyLysProSerProLysAlaAlaSerSerProSerAsnProAla 1114
2747 GCTGGGC 2753
1114 lAlaLeuProValAlaSerAspSerSerProMetGlySerLysThrLysGlu 1130
2754GCTGCCTAATCTAGTGTGCTGCTCAACAGGGCA 2792
1131 ThrAspSerProSerThrProGlyLysAspGlnArgSerMetLleLeuAr 1147
2793 TGGAAATTAAAGCATGGCTCCAG 2815
1147 gSerArgThrLysThrGlnLlePheHisSerLysArgArgArgProS 1164
2816 AAGTTACAGAATCTCTTGGGATCTTCTCGCAAACTTCTCCAGCCAA 2865
1164 erGluGlyArgLeuProAsnCysArgAlaThrLys 1175
2866 AAGCAGCGGTCTCCAGGAATGCCAGTCAAAAAGGTATGGCGCGCC 2915
1176 ...LysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValSerSe 1191
2916 ATGAGACTGAGGACATGGACTAGCTGAGGCTAC 2949
1191 rSerProGlnLysGluGlyArgValSerGlnArgAlaArgValProLysP 1208
2950 ACAGTCATCCAAACCTGCTAGTGTATGCTGAGACTTC 2987
1208 roGlyAlaGlySerLysLeuSerAspArgProLeuHisAlaLeuLysArg 1224
2988 CAGCCAGGAGGATCATCTCTCAAAACCCCTTAATCATGAGGAGCGT 3037
1225 LysSerAlaPheMetAlaProValProThrLysLysArgAsnLeuVal 1241
3038 GTTCGTCTTTTATCTCTCCATCCAGTAAAGACAGACAGTCAAGATGT 3087
1241 uArgHisGlySerSerSer 1248
3088 AAGAAGACGTAGCACTGAAGATAAAGGTGCTCTCTCACTCATCAAAAG 3137
1249SerAsnAlaSer 1252
3138 AAGCGCTGATTAAGCATTCAATCTCTGCCCCTTCTCAGACTCAG 3187
1253 AlaMetGlyGluMetGlyArgArgGlyLeuArg...ValProProPr 1268
3488 GATATCAATCTATCCCTAAGAGATTTCTCCAAAGACCTTCCAAAGTCC 3237
1268 sSerSerArgGlyCysLeuLeuProArgLysProSerProArgAlaM 1285
3238 AGATAGTAGAAACTGCCCTGTGTACCTCTCAAGCCCT 3277
1285 eAlaSerLeuProGlnSerSerHisProArgArgProProPheLeuPro 1301
3277 3277
1302 GlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysVal 1318
3278GCTAAGACCAAAATACT 3294
1318 uProProArgLysGlyArgGlyLeuLysLeuGluAlaLleValGlnLysI 1335

```

|||||
3295 GCCCCACGAGAGAGAGGAGTTGAAATGGAAGCTATAGTTCAGAGAA 3344
1335 IethiserProserleuLysPheAlaCysLysAlaProglLysaser 1351
|||||
3345 TTAAATCCCCAAATATAGAGAGAGCGCATCTCGAAACAGTCGGAGAGCT 3394
1352 ProglLysnProLysenProserLysenLysasprLysglLysleu 1368
|||||
3395 GGGGAGAGACAGGTTACG.....CTTGATGATATACTGCTTTGAA 3435
1368 SGLYALA.....GlyLysenProvalGlyAlaGluGlyLeu 1382
|||||
3436 GAGTGTCTCTCTGAGAGTGGAGCT...GTTCGTCTTACAGAGAGTCA 3482
1382 AlaenValGlyThrGlyLysLeuProThrSerGlyAlaAspProleu 1398
|||||
3483 TAGAGAGAGAGAGAGAGTGAAGTGA.....GCTTGAGACTTA 3517
1399 CysArgasnProThrAsnArgSerleu.....LysGlyLysleuMetas 1413
|||||
3518 GTC...AGTCCAGCAACCAAGAGATTCACGTAGAGAAACCTCTCCAG 3564
1413 nserLys.....LysLeuSerSerThrAspCysPheLysThrGlu 1426
|||||
3565 GTCTTCAGAGAGAGTGGCGGTGGACAGCTGATGACAAAGTGAAGACAGA 3614
1427 ....AlaPheThrSerProGluAlaLeuGlnProGlyGly..... 1438
|||||
3615 CACATGCGAGAAACAGTTACTGCGGAAAGAGAACCCCTGGTGCCAGACA 3664
1439 ThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAla...GlyAl 1454
|||||
3665 TCCACACCTCCACAGAGAGCTGTGATACCAAGAGAGAGACAGATGGTTC 3714
1454 aHisGlyLeuSerLysGlyProLeuGlnLysArgProGlyLeuGlyPro 1471
|||||
3715 CCGGGT.....GGAACAG 3728
1471 laLeuLeuLeuThrProArgAspAlaGalaSerGlyThrGlnGly..... 1485
|||||
3729 CACCTTATCTTCCAGAGCTCAAGAAATGATACCTCAGTGGGCTATATG 3778
1486 AlaSerGluAspAsnSerGlyGlyGlyLysLys..... 1497
|||||
3779 GCCCTGAGCAACCCAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3828
1498 .....ProLysMetGluGlnLeuGlyProAlaSerGlnProPro.... 1510
|||||
3829 GATTTCACCGAAGAGAGAGGTTTCCCTCCAAAGGATATTTCCCATCAG 3878
1511 .....GluGlyArgProCysGlnProGlnThrArgAlaGlnLysGln 1524
|||||
3879 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3928
1525 .....ProGlyAla 1527
|||||
3929 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAGAGTGC 3978
1527 sThrAsnLysSerSerLysLysArgLysArgLeuThrArgGlyArg 1544
|||||
3979 TGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4028
1544 laLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArg 1560
|||||
4029 GGAAG.....CTGGGGCCAGCCGAGAGAGAGAGAGAGAGAGAGAG 4063
1561 GlnGlnGlnValLeuProLeuAspProAlaGlnProGlnLysLeu 1577
|||||
4064 AARCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4113
1577 sThrLysSerCysLysArgLeuArgSerArgSerArgThrProAla 1594
|||||

```

```

4114 AATGCCACCCAG...CCACTGATTAATAAGTATGTCAGAGAGAGAGCTT 4160
1594 hSerProPheValArgValGluLysArgAspAlaPheThrThrLeu 1610
|||||
4161 TTACCTTACATCCATGATGAAATATAGTGAATGAGTGAAGCGGTTGT 4210
1611 ThrValAlaAsnSerProGlyLysP.....AlaPr 1620
|||||
4211 ACATATCATCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260
1520 oLysProHisArgLysProSerSerSerAlaSerSerSerSerSers 1637
|||||
4261 GAAGGTCAAGAGTCACTACCTCCACCTACAGAGAGAGAGAGAGAGAG 4309
1637 erSerPheSerLeuAspAlaValAlaGlyAlaSerLeuAlaThrLeuProGly 1653
4309 ..... 4309
1654 GlySerThrLeuGlnProArgProSerLeuProLysSerThrMetH 1670
4310 .....GCGTCCCGGCTCGTCTTATGCT 4335
1670 sleuGlyProValValSerLysAlaLeuSerThrSerCysLeuValCys 1687
|||||
4336 CGAGGAGAGCTGTGTGACAGAGCTTCGTTATGGGACAGCTGTTGCT 4385
1687 yLeuGlnGlnAsnProAlaAsnPheLysAspLeuLysLysLeuGly 1703
|||||
4386 GTCTGTGTGCAAGTGGGCGCATTCACGAGAGAGAGAGAGAGAGAGAG 4435
1704 ProTyrThrProGlnHis.....CysLeuProLysLysPro.. 1716
|||||
4436 CTTTTTATCCCAAGATTTATGACAGCAGCTCTCCGAGAGAGAGAGAGAG 4485
1717 .....LysLeuLysGlnLysValArgProGluGly 1727
4486 TAAGAGGCGCACAGAAATGACAGAGAGAGAGAGAGAGAGAGAGAGAG 4528
1727 hrCysGluGluAlaSerLeuProLeuGluArgThrLeuLysGlyProG 1743
|||||
4529 .....AAAAGTCTTATGAGGCTCCAGAGAGAGAGAGAGAGAGAGAG 4573
1743 lucYsaAlaAlaAlaAlaThrAlaGlyLysProPro 1754
4574 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4608
seq_name: gb_sts:AF021118
seq_documentation_block:
LOCUS AF021118 495 bp DNA STS 03-APR-1998
DEFINITION Homo sapiens trilineoleotide repeat ctg-37, sequence tagged site.
ACCESSION AF021118
VERSION AF021118.1 GI:301591
KEYWORDS STS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 495)
Phillibert,R.A., Horelli-Kuitunen,N., Robb,A.S., Lee,Y.H.,
Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,
Palotie,A. and Gims,B.I.
The characterization and sequence analysis of thirty CTC-repeat
containing genomic cosmid clones
Eur. J. Hum. Genet. 6 (1), 89-94 (1998)
JOURNAL
MEDLINE
98454304
REFERENCE
2 (bases 1 to 495)
Phillibert,R.A., Horelli-Kuitunen,N., Robb,A.S., Lee,Y.H.,
Long,R.T., Damschroder-Williams,P., Martin,B.M., Brennan,M.B.,
Palotie,A. and Gims,B.I.
Direct submission
JOURNAL
Submitted (28-AUG-1997) Clinical Neuroscience Branch, National
Institutes of Mental Health, 49 Convent Drive Rm B1E16, Bethesda,

```



```

FEATURES
  source
    MD 20892, USA
    Location/Qualifiers
      1..495
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /maposome="17"
        /map="17q11.2"
        /note="primers to amplify repeat: forward:
        AAGAGGCGCAGGGTGACCCAG; reverse:
        AGACATCTCTTGAACAGGGTGG"
BASE COUNT 124 a 165 c 135 g 67 t 4 others
ORIGIN
  alignment_scores:
    Quality: 625.00 Length: 156
    Ratio: 4.340 Gaps: 2
    Percent Similarity: 92.308 Percent identity: 82.051
  alignment_block:
    US-09-508-821B-6 x AF021118
  Align seg 1/1 to: AF021118 from: 1 to: 495
    1124 MetGlySerLysThrLysGlu...ThrAspSerProSerThrProGlyLys 1139
    1 ATGGGGCTCCAGACACAGGAGACAGACTCACCAGCAGCGCTGCGAAG 50
    1140 AspGlnArgSer MetLeuArgSerArgThrLysThrGlnGluLeuP 1156
    51 GACCAGCGCTCCATTGATTCCTGGTCAAGCACCACCAACCCAGGAGATCT 100
    1156 heHisSerLysArgArgProSerGluGlyArgLeuProAsnCysArg 1172
    101 TCCACTCCAAAGCGCGAGGGCCCTCTGAGGGCGGGCTCCCAACTCCGT 150
    1173 AlaThrLysLysLeuLeuAsnSerHisLeuProAlaThrPheLysVa 1189
    151 TCCACCAAGAGCTCTCTGACACACACACTTTTCGCCACATTCAGGT 200
    1189 lSerSerProGlnLysGluGlyA-gValSerGlnArgAlaArgVaIP 1206
    201 CTCACACAGCCCCAGAGAGGCGGAGTGAACACAGCGGCGGAGGTCC 250
    1206 rolyProGlyAlaGlySerLysLeuSerAspArgProLeuHisAlaLeu 1222
    251 CCAAACTGGTCKAGGCGAGGCTCTCTGACCGGCCCTCCATGGCGTC 300
    1223 LysArgLysSerAlaPheMetAlaProValProThrLysLysArgAsnLe 1239
    301 AAAGAGAGTGGCGCTCATGGCGCGGTCCGCCACCAAGAGCGGAACCT 350
    1239 uValLeuArgHisGlySerSerSerSerSerAsnAlaSerAlaMetGlyG 1256
    351 GGTCTTGGAGCGAGCGACGACGACGACGACGACGACGACGACGACG 400
    1256 LuMetGlyArgArgGlyLeuArgValProProProSerSerArgGly 1272
    401 AGATGGGAAGGAGGAGGCGCTGANGGTTCGCCGCCCTCTTCAAGAGGA 450
    1273 CysLeuLeuProArg 1277
    451 TGGTNTCTCCAGA 465
  seq_name: gb_sts:G09710
  seq_documentation_block:
    LOCUS G09710 542 bp DNA STS
    DEFINITION human STS CHLC.GCT10D04.P16019 clone GCT10D04.
    ACCESSION G09710
    VERSION G09710.1 GI:941559
    KEYWORDS STS sequence; primer; sequence tagged site.
    human vector-pJUCP1 host-E.coli dut-tung+ (DH10B) Marker Selected

```

ORGANISM

genomic DNA prepared from XY individual of French nationality.
Homo sapiens
Eukaryota; mitochondrion eukaryotes; Metazoa; Chordata;
Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Choanata;
Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
Catarrhini; Hominoidea; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 542)
Murray J., Sheffield V., Weber, J.L., Duyk, G. and Buetow, R.H.
Cooperative Human Linkage Center
Unpublished (1995)
Synonyms: GCT10D04, CHLC.GCT10D04.T15491
Contact: Dr. Jeffrey C. Murray

UofI

The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jeff-murray@uiowa.edu

Primer A: CTGCGGTAGTGTGATGACT

Primer B: ATCTTCATGCTACAGTCG

STS size: 155

PCR Profile:

denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degrees C

Protocol:

Template: 30ng genomic DNA
Primer: each 1.5 pmole
dNTPs: each 200 uM
Taq Polymerase: 0.3 units
Total Vol: 10 ul

Buffer:

MgCl2: 1.5mM
KCl: 50mM
Tris: 10mM
pH: 8.3.

FEATURES

source

Location/Qualifiers

1..542

/organism="Homo sapiens"

26..180

26..45

complement(161..180)

82 a 117 c 168 g 131 t 44 others

primer_bind

primer_bind

BASE COUNT

ORIGIN

alignment_scores:

Quality: 595.50 Length: 151

Ratio: 4.546 Gaps: 5

Percent Similarity: 86.755 Percent Identity: 82.119

alignment_block:

US-09-508-821B-6 x G09710/rev

Align seg 1/1 to reverse of: G09710 from: 1 to: 542

179 ProGlnGlnProLeuAlaArgPro.LysLeuGlnArgGlnLysLeu...G 194

450 CCAGNAGGCCCTGGNATACCCNAAGCTNCCAAAGGNAANNTGAA 401

194 InAsnAspIleAlaSerProLeuProPhe...ProGlnGlyThrHisPhe 209

400 GAAAGANATTGCTNNCNTNTGCTTTCCTTCCCGGAGGACCACTTT 351

210 ProGlnHisSerGlnSer...PheProThrSerSerThrTyrSerSer 225

350 CCTCAGNATTCACCAAGTCTTCCCGACCTCTCCACCTACTCTCTCTC 301

225 rVal.GlnGlyGlyGlnGlyAlaHisSerTyrLysSerCysThrAla 241

[illegible]

[illegible]

736 tSerAlaSPSerAlaSPProPheAlaTrpProGlnGlnAlaSPLeuGlyA 753
::: |||||::: ||::: |||||:::
130548 GCTAGCCAAACCGACGACGACCGCCAGCCAGCCAGC..... 130512

753 SPAlaSerProAlaGTrpGlyLeuAlaSPGlnGlnLeuThrLysGlyLeu 769
|||||::: |||||:::
130511CACCCAGCC..... 130503

770 GlnGlnGlyGlyLeuAlaSerAspGlyLysSerLysGlyAspThrHisGln 786
::: |||||::: |||||::: |||||:::
130502 AGCCAGACCCAGCCAGCCAGCAATCCAGCCAGCCAGGACGACGCCAGCCAGG 130453

786 ValaSerAlaCysLeuGlyPheGlnGlnGlnGlnAlaSPProGlnGlyLysVal 803
::: |||||::: |||||::: |||||::: |||||:::
130452 CAGCAGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGC... 130407

803 AlaAlaSerLeuProGlnGlyAspPheLysGlnGlnGlnValaGlyValLys 819
|||||::: |||||:::
130406 ..CCAGACCCGCCA.....GAC 130392

820 GlnGlnGlnGlyGlyLeuGlnGlnCysProGlnValaAlaLysAlaSPAr 836
::: |||||::: |||||::: |||||:::
130391 AGCCAGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130342

836 GTrpLeuGlnAlaSPSerAlaHisCysLysCysSerThrAlaSPPheGlyAspL 853
::: |||||::: |||||::: |||||:::
130341 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130303

853 euProLeuLeuProProThrSerArgLysGlnaSPLeuGlnaGlnGln 869
::: |||||::: |||||::: |||||::: |||||:::
130302 ..CCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130255

870 GlnTrpSerSerLeuCysGlnLeuLeuGlnGlySerProGlnGln..... 883
::: |||||::: |||||::: |||||::: |||||:::
130254 AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130205

884ArgProGlnLysLeuGln..... 888
::: |||||::: |||||:::
130204 GCCAGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130155

889AspProLeuSerProLysAlaPro 896
::: |||||::: |||||::: |||||:::
130154 CAGCAGCCAGCCAGCCAGCCAGCCAGCTAAGGCAAGCCAGCCAGCCAGCCAGCCAG 130105

897 LeuLleCysThrLysGlnGlnValaGlnGlnValaLleuAspSerLysAlaGln 913
::: |||||::: |||||::: |||||::: |||||:::
130104 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130055

913 YTrpGlySerProCysHisLeuSerGlyGlnSerValLleuLeuLeuLysPro 930
::: |||||::: |||||::: |||||::: |||||:::
130074 T...GGCAGCCCAACCCAGCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 130014

930 roThrValaGlyThrGlnSerLysValaGlnSerTrpPheGlnSerSerLeu 946
::: |||||::: |||||::: |||||::: |||||:::
130013 GCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 129964

947 SerHisMetLysTrpGlyGlnGlnGlyProAspGlyGlnLys..... 960
::: |||||::: |||||::: |||||::: |||||:::
129963 CAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 129920

961AlaProG 963
::: |||||::: |||||:::
129919 AAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 129870

963 LysAspSerThrTrpSerAlaSerLeuAlaGlnLysProAsnLysPro 979
::: |||||::: |||||::: |||||::: |||||:::
129865 CCAAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 129820

980 AlaValProGlnAlaProLleAlaLysLysLysLysLysLysLysLysLysLys 996
::: |||||::: |||||::: |||||::: |||||:::
129813 AACCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 129770

[illegible]

|||||
490 ATGCACTCTTTTCGAGAAAGGTGTGTTTCCATGCGCAACACAGAACTA 539
17 rGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyrArgGlnPro 34
|||||
540 CCAGCAGACCTCCGAGAAACATCATCGGCTAGAGATTACAGCAGCGCA 589
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLysAspTyr 50
|||||
590 GTGAGCGCGGGTAACTGCTGAGCGGCGGCGGCTGCTGCGCAAGACTAT 639
51 TyrAsnProGlnProTyrProSerTyrGlnGlyValAlaGlyThrProSe 67
|||||
640 TATAACCCGACAGCTTACCGAGCTATGAGGCTGAGGCTGGCAGCGCTC 689
67 rGlyThrAlaAlaAlaValAlaAlaLysPheTyrHisArgGlySerLys 84
|||||
690 TGGCAGCTGCAGCGCGGTGGCCCGCCGACAAAGTACCCAGGAGCAAGG 739
84 IaLeuProThrGlnGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTyr 100
|||||
740 CCTTGGCCACACAGCAGGCTGCAGGAGGCGGCTTCCCTGGCTAC 789
101 GlyValGlnAspSerProTyrProGlyArgTyrAlaGlyGluGlu 117
|||||
790 GCGCTCCAGCAGCAGCGCTTACCCAGCGCGCTATGCTGCTGAGGAGAG 839
117 rLeuGlnAlaTTPGlyValProGlnProProProGlnProGlnProL 134
|||||
840 CTTTCAGGCTTGGGGGCGCCAGCAGCCACCCCGCCAGCGCGCGCC 889
134 euProAlaGlyValAlaLysTyrAspGluAsnLeuMetLysThrAla 150
890 TACCTGAGGCGGTGGCCAAATATGATGAGAGCTTGAATAAAGACAGCA 939
151 ValProProSerArgGlnTyrAlaGlnGlnGlyAlaGlnValProPhe 167
|||||
940 GTGCCCCCAGCAGCAGTATGAGAGCGGCGCCAGGTGCCCTTTTCG 989
167 gThrHisSerLeuHisValGlnGlnProProProGlnGlnProLeu 184
990 GACTCACTCCCTGCAGCTCCAGCAGCAGCGCGCGCCAGCAGCGCTGG 1039
184 IaTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerPro 200
1040 CATACCCCAAGTCCAAAGCAGAGCTGCAGAGCAGCAFTGCTCCCTCC 1089
201 LeuProPheProGlnGlyThrHisPheProGlnHisSerGlnSerPhe 217
1090 CTGCCCTTCCCGCAGGGTACCCTTCTCAGCATTCGCCAGTCTTCC 1139
217 oThrSerSerThrTyrSerSerSerValGlnGlnGlyGlnGlyAlaH 234
1140 CAGCTCTCCACCTACTCTCTCTCTCCAGGGTGGTGGCGAGGGGGCC 1189
234 sSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArgPro 250
1190 ACTCCTATAGAGTTGCACAGCAGCCGACTGCCAGCGCCCATGACAGCG 1239
251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuH 267
1240 CTGACTGCCAGCTCCAGCTTGGCGCGCGGCGGCGGCTCCAGAACTTCA 1289
267 sAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGln 284
1290 TGGCTACCATGCGGGCGGCTCAGCTATGACCAGCAGCAGCAGCAGC 1339
284 lngGlnGlnGlnGlnGlnGlnAlaLeuGlnInSerArgHisHisAlaGln 300
1340 AGCAGCAGCAGCAGCAGCAGCAAGCGCTTTCAGAGCGGCGGCACTGCCAG 1389
301 GluThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlyGln 317
|||||

1390 GAACCCCTCCATTACAAAACCTCGCCAGATATCAGCACTACGGGCGAGCA 1439
317 nGlyGlnGlyTyrCysGlnProAspAlaAlaValArgThrProGlnGln 334
1440 AGGCCAGGCTACTGCGCGGAGCGGAGCGCTCGGACCCCGCAGAGCAGT 1489
334 yTyrGlnThrPheSerProSerSerSerHisSerProAlaArgSerVal 350
1490 ACACAGAGCTTCCAGCCCGAGCTCCAGCAGCTCAGCCCGCGGCTCGGT 1539
351 GlyArgSerProSerTyrSerThrProSerProLeuMetProAsnLe 367
1540 GSCCGCTCAGCTTCTTACAGTTCACACCGCTGCGCGCTGATGCCAACT 1589
367 uGluAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaPhePro 384
1590 GAGAACTTCCCTACAGCGCAGCGCGCTCAGCAGCGGCGCTTCCCG 1639
384 lagIleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400
1640 CAGGATCACTGACACAGCAGCTTCAAGCCCTGCTCAATCCCTCCCA 1689
401 ThrAspAlaThrSerSerValAspThrGlnAlaGlyAsnCysLysProLe 417
1690 ACGGATGCCAGCAGCTCTGTGACACCCAGGCTGGCACTGCAAGCGCT 1739
417 uGlnLysAspLysLeuProGluAsnLeuLeuSerAspLeuSerLeuGln 434
1740 TCAGAGGAGCAGCTCCCTGAGAACCTCTGTCGGATCTCAGCTGCGAGA 1789
434 erLeuThrAlaLeuThrLeuGlnValGluAsnIleSerAsnThrValGln 450
1790 GCCTCAGCGCTGACCTTACAGGTGGAGAACATCTCCAAACCGTCCAG 1839
451 GlnLeuLeuLeuSerLysAlaAlaValProGlnLysGlyValLysAs 467
1840 CAGCTCTCTCTCAAGGCTGTGCGCGAGAAAGGTGCAAGAA 1889
467 nLeuValSerArgThrProGlnHisLysSerGlnHisCysSerProG 484
1890 CCTCGTCCAGGACCCAGCAGCAGTAAAGCCAGCAGCTGCAGCGCCG 1939
484 luGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSerGluPro 500
1940 AAGGAGCGGCTACTCAGCGGAGCGCCGAGCAGCAGCTGTCAGCGCG 1989
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGlnAlaAspTy 517
1990 CCGAGCAGCAGCCACAGTCCAGCATCGGAGCGCGCAGGCGCGACTA 2039
517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCysAsnG 534
2040 CCGTGGCGGCTCCAGAGCCACTGGAGCGCAGCTTCTTACTGCAACC 2089
534 lnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550
2090 AGGCCGCTGGCAGCGCTGCCAGGCTCAACAGCACTCGAGGCCAAGCC 2139
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrLy 567
2140 GAGTCCGCTGCCACTTCTGTGACCTCTCTGAGCAGATGTCACCAA 2189
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584
2190 ATCTGAGGAGCTTCCAGAGCCTACAGGAGCTTCCGCTCGACAGCT 2239
584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuLeuSer 600
2240 TCTCCAGTTCGTGGCGGCTGAGCGGAGCTGTCCGCGGCTGCTGCTCAGC 2289
601 AlaLeuAlaGlnGluAspLeuAlaSerGlnIleLeuGlyLeuGlnGlu 617
2290 GCCTCGCAGGAGGAGCCCTGGCTCCGAGATCTCGGCTGCAGAGAGC 2339

617 attleGlyGluValAspLysAlaTrpAlaGluAlaProSerLeuValL 634
|||||
2340 CATCGTGAAGAGCCGACAAAGCTGGCTGAGACACCCAGCTGTGCTA 2389
634 yAspSerLeuLysProPheSerLeuLysHisSerAlaCysLeu 650
|||||
2390 AGAGACAGCAGACCAACCTTCTGCTGAGAGAACCCAGCCCTGCTG 2439
651 AspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAlaLe 667
|||||
2440 GACTCTGTGGCAGAGAGTGTGGCCCGGCTGGGAGCCGAGAGGCCCT 2489
667 uProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684
|||||
2490 GCCCGACTCTTGTGACCTGTGACAGAGGGGCGCAATGCCAAGACTTCAACC 2539
684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700
|||||
2540 CAGGGCTGTTTGAAGACCTTCCGTGGCTTGTGCTGACGCTGACCCCAA 2589
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717
|||||
2590 AAGACACTGTGTCTCTCTCTTGTGTACCAAGCCCACTTGGGGTTCC 2639
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThra 734
|||||
2640 TGCTCCAGACCCCACTACAGACACTTGTGACGTTCCTCCGAGACACACG 2689
734 lAlaAspSerAlaAspSerAlaAsnProPheAlaTrpProGluGluAsn 750
|||||
2690 CTGCCACCTCAGGAGACGCCACCCCTTGTCTGCTGCGCCAGAGGAAAC 2739
751 leuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrLys 767
|||||
2740 CTGGGGATCTCTTCCAGGTGGGATGTGACCCCTGGCGACTACCA 2789
767 sGlyLeuGluGlnGlyGlyLysAlaSerAspGlyLysSerLysGlyAsp 784
|||||
2790 GGGCTGTGACAGGAGGTGGAGGCTGAGTGGCACTACAGCAAGAGGAGCA 2839
784 hHisGluAlaSerAlaCysLeuGlyPheGlnGluGluLysProGly 800
|||||
2840 CCCATGAGGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2889
801 GlnLysValAlaSerLeuProGlyAspPheLysGlnGluGluValGly 817
|||||
2890 GAGAGGTGGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2939
817 yValLysGlnGluAlaGlyGlyLeuLeuGlnCysProGluValAlaLys 834
|||||
2940 GGTGAAGAGAGAGCAGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2989
834 lAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850
|||||
2990 CTGACCGGTGTGTGAGAGACGCCACTCTCTTCCACCCGCACTCTC 3039
851 GlyAspLeuProLeuLeuProProThrSerArgLysGlnLysLeuGluAla 867
|||||
3040 GGGGACCTCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3089
867 agLugLugLysSerSerLeuCysGlnLeuLeuGlySerProGluGluAla 884
|||||
3090 TGAAGAGAGTACTCTCTCTATGTAGTCTCTGAGAGCCGAGAGCA 3139
884 rGProGlyLeuGlnAspProLeuSerProLysAlaProLeuLysCysThr 900
|||||
3140 GGCCTGTGATGAGAGCCGCTGTCTCCAGAGCCCACTCACTGCAAC 3189
901 LysGlnGluValGlnGluValLeuAspSerLysAlaGlyTrpLysThr 917
|||||
3190 AAGAGAGAGGTGTGAGAGGTCTGTGACTCCAAAGCCGCTGGGGCTCTCC 3239

917 cCysHisLeuSerGlyGluSerValIleLeuLeuGlyProThrValGlyTr 934
|||||
3240 GTGCCACCTCTCAGAGGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3289
934 hrgLysLysValGlnSerTrpPheGluLysSerLeuSerHisMetLys 950
|||||
3290 CCGAGTCAAGAGTCCAGAGCTGTTGAGTCTCTCTCTCTCTCTCTCTCTCT 3339
951 ProGlyGlnGluGlyProAspGlyGluArgAlaProLysAspSerThr 967
|||||
3340 CCAGGTGAAGAGGGCTGATGTGGGAGCAGAGTCCAGAGGATTCACAC 3389
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluAla 984
|||||
3390 CTCGAGACCTCTCTGCGCCAGAGCCCAACAGCTCTCTCTCTCTCTCTCT 3439
984 lAProIleAlaLysLysGluProValProArgLysSerLeuArgSer 1000
|||||
3440 CGCCCTTCGCAAGAAAGAGCTGTGCAAGGGGCAAAAGCTTACAGAAC 3489
1001 ArgArgValHisArgLysLeuProGluAlaGluAspSerProCysArgL 1017
|||||
3490 CGTGGGTGTGACCGGGGCTGCGCGAGCCGAGAGCTCCCAATGCAAGG 3539
1017 aProValLeuProLysAspLeuLeuProGluSerCysThrGlyProP 1034
|||||
3540 ACCAGTGTGCGCAAGAGCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCT 3589
1034 roGlnGlyGlnMetGlnGlyAlaGlyAlaProGlyValArgLysAlaSerGlu 1050
|||||
3590 CCGAGGAGCAGATGAAGGGGCTGAGGCCCAAGCGGGGGCTCGGAA 3639
1051 GlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProAr 1067
|||||
3640 GGGCTCCCGAGAGTGTACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3689
1067 gThrProGlyProProGlyLeuThrThrThrProAlaProAspLysL 1084
|||||
3690 CAGCGCCGAGCCCGAGGCTGTGACACACCCCTGCAACCCCAAGCAAC 3739
1084 euGlyLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100
|||||
3740 TGGGGGCAAGCAGAGAGCGCTTCAAGTGGGCAAGCGGGGTGAGAG 3789
1101 pSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117
|||||
3790 CCGTCAACCAAGGCTGCTCTCCAGCCCAACCGGCGGCTGCTGCTGT 3839
1117 lAlaSerAspSerSerProMetGlySerLysThrLysGluThrAspSerP 1134
|||||
3840 GGCTCCGACAGACCCGATGGGTCTCAAGACCAAGAGACAGACTCAC 3889
1134 rSerThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThr 1150
|||||
3890 CCAGACGCTGGCAAGAGACAGCGCTCCATATCTTCGGTCAACGAC 3939
1151 LysThrGlnGluIlePheHisSerLysArgArgArgProSerGlnLys 1167
|||||
3940 AAAACCGAGAGATCTCCACTCCAAACCGGCGAGGCTCTGTGAGGCG 3989
1167 gLeuProAsnCysArgAlaThrLysLysLeuLeuAspAsnSerHisLeuP 1184
|||||
3990 GCTCCCAACGTGCGTCCACCAAGAGCTCTCTCAACAGCCACTGTG 4039
1184 rAlaThrPheLysValSerSerProGlnLysGlnLysValSer 1200
|||||
4040 CCGCAGCTTCAAGGTCTCCAGAGCCCGCAAGAGAGGAGGAGGTGAG 4089
1201 GlnArgAlaArgValProLysProGlyValGlySerLysLeuSerAspAr 1217
|||||
4090 CAGCGGCAAGGTCGCCAAACCTGTGCAAGGAGGAGAACTCTCTGACG 4139
1217 pProLeuHisAlaLeuLysArgLysSerAlaPheMetAlaProValProT 1234

5040 CCAGACAGGGGACAGAAAGCCAGGCGCCACCACTACAGCAGCTATT 5089
1534 erLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSerSer 1550
5090 CCAAGCGGAGGCGCTCACTCGGGCGGGCCAAAGAACACCACTCTTCA 5139
1551 ProCysLysGlyArgAlaLysArgArgGlnGlnValLeuProLeu 1567
5140 CCCTGAAGGGCGTGGCCAAAGCGACGACAGCAGCAGGCTGCGCCCT 5189
1567 uapProAlaGluProGlnLeuLeuLysLysLysLysLysLysLys 1584
5190 GGATCCCGCAGAGCGCTGAATCCGCTCAAGTACATTTCTCTTCAAGC 5239
1584 rgLeuArgSerArgThrProAlaPheSerProPheValArgVal 1600
5240 GCGTCAGGTGACAGCGGAGCGGCGCTTCTACCGCTTGGTGGGGTG 5289
1601 GluLysArgAlaPheThrThrIleCysThrValValAsnSerProGln 1617
5290 GAGAAGCGGAGCGCTTCCACCCATGACCTGTGTCACTCCCTCTTCT 5339
1617 yaspAlaProLysProHisArgLysProSerSerSerSerSerSerS 1634
5340 AGTGGCGCCCAAGCCACAGAGAGGCTTCTCTCTCTCTCTCTCTCT 5389
1634 erSerSerSerSerPheSerLeuAspAlaAlaLysLysLysLysLys 1650
5390 CATCT 5439
1651 LeuProGlyGlySerIleLeuGlnProArgProSerLeuProLeuSer 1667
5440 CTCTCTGAGGCTCCATCTCTGCGCGCGGCGCTCTCTCTCTCTCTCT 5489
1667 rThrMethHisLeuGlyProValValSerLysAlaLeuSerThrSer 1684
5490 CAGATGCGCTTGGGCGCTGTGTTTCCAGGGCGCTGAGTACCTCTTGC 5539
1684 euValCysCysLeuCysGlnAsnProAlaAsnPheLysAspLeuGly 1700
5540 TTCTTTGCTGCTCTGCAAAACCCGCGCACTTCAAGGACCTTGGGAC 5589
1701 LeuCysGlyProTyrTyrProGluHisCysLeuProLysLysLysPro 1717
5590 CTCTGGGCGCTTACTTCTCTGACACTGCTCTCTCTCTCTCTCTCTCT 5639
1717 sLeuLysGlyLysValArgProGluGlyThrCysGluGluAlaSerLeu 1734
5640 ACTCAGGAGAGGTGGCGGCGAGAGGACCTGTGAGGAGGCTCTCTGC 5689
1734 roLeuGluArgThrLeuLysGlyProGluCysAlaAlaAlaAlaThrAla 1750
5690 CGCTTGAGAGAACACTCAAGGTCCCGAGTGGCAGCTGCGCCCTGCTGC 5739
1751 GlyLysProProArg 1755
5740 GGAAGCGCGCGCGG 5754

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ88975

seq_documentation_block:

ID AAZ88975 standard; cDNA; 5395 BP.

XX

AC AAZ88975;

XX

DT 26-MAY-2000 (first entry)

XX

DE Human HSGT1 cDNA.

XX

XX HSGT1; human; chromosome 17; short arm; neuroprotection; diagnosis;

KW Smith Magen's Syndrome; multiple congenital anomaly; ss.

XX

OS Homo sapiens.

XX

1410 GCGCCCTCATGCGCTCAAAAGGAAGTGGCTTCAATGGCGCGGTCCCA 4189
1234 hrLysLysArgAsnLeuValLeuArgHisGlySerSerSerSerSerAsn 1250
4190 CCAAGAAGCGAACCTGGTCTTGGCGGACGCGCAGCAGCAGCAGCAAC 4239
1251 AlaSerAlaMetGlyGluMetGlyArgArgArgGlyLeuArgValProPr 1267
4240 GCGATGCAATGGGGAGATGGAGAGGAGGAGGAGGCGCTGAGGGTCCGCC 4289
1267 oProSerSerArgGlyCysLeuProArgLysProSerProProArgA 1284
4290 ACCCTCTTCAAGAGATGCTTCTCCCAAGAAAGCCACGCCCCACAGGG 4339
1284 laMetAlaSerLeuProGlnSerSerHisProArgArgProProPheLeu 1300
4340 CAATGGCGAGCTGCCACAAAGCTCCACCCCGGAGAGCCCACTCTCTG 4389
1301 ProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysVa 1317
4390 CTTCAAGCTCGCTCTGGGAGCTTCCAGGGGCGCATGAGACCAAGGT 4439
1317 lLeuProProArgLysGlyArgGlyLeuLysLeuGluAlaIleValGlnL 1334
4440 GCTGCCACACCCCGAAGGCGGGGCGCTGAGCTGGAGGCGCATCGTGCAG 4489
1334 ysileThrSerProSerLeuLysLysPheAlaCysLysLysAlaProGlyAla 1350
4490 AGATCAGCTCGCCAGCTCAGAGTTCGCTGATGTAAGGCGCGAGGGGCC 4539
1351 SerProGlyAsnProLeuSerProSerLeuSerAspLysArgGlyLe 1367
4540 TCTCTCTGTAATCTCTGAGCGCATCCCTTCCGAAAGACCGTGGGCT 4589
1367 uLysGlyAlaGlySerProValGlyValGluGluGlyLeuValAsnV 1384
4590 CAAGGGTCTGGGGGCGCGCGGCTGGGGGTGGAAGAAGGCTGGTAAAG 4639
1384 alGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeuCysArg 1400
4640 TGGGCGCGGCGAGAGCTCCCACTCTCTGGGCTGATCGTATGCGAGA 4689
1401 AsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLysLe 1417
4690 AATCCACCAACAGATCTTAAAGGCAAACTCATGACAGTAGAAGACT 4739
1417 userSerThrAspCysPheLysThrGluAlaPheThrSerProGluAla 1434
4740 GTCTTCTACTGACTGTTTCAAAACCGAGGCGCTTCACTATCCCGGAGGCC 4789
1434 euGlnProGlyThrAlaLeuAlaProLysLysArgSerArgLysGly 1450
4790 TGACGCTGGGGGAGTGGCTGCGCGCTTAAAGAGGAGGAGCGGAAAGGC 4839
1451 ArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgProTy 1467
4840 CGGCGAGGGCGGCGGCTGCTCTCAAGGCGCGCTGGAGAGCGGCCCTTA 4889
1467 rLeuGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGlyThrG 1484
4890 TCTTGGCGCGGCTCTGCTCTCTGACTCCCGGAGAGCGGCGGCGGCGCAC 4939
1484 lngLysAlaSerGluAspAsnSerGlyGlyGlyLysLysLysProLysMet 1500
4940 AAGGGCGGAGTGGGAGCACTCTGGTGGAGGAGGCAAGAGCAAGAGATG 4989
1501 GluGluLeuGlyProAlaSerGlnProProGluGlyArgProCysGlnPr 1517
4990 GAGGAGCTGGGCGCTCTCCAGCGCGGAGGAGGCGGCGGCGGCGGCGG 5039
1517 oGlnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrS 1534

XX MO200008143-A2.
 XX 17-FEB-2000.
 XX
 XX 06-AUG-1999; 99WO-DE02479.
 XX 07-AUG-1998; 98DE-1035910.
 XX
 XX (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
 XX (PLAC) MAX-PLANCK-INSTR MOLEKULARE GENETIK.
 XX
 XX Seranski P, Pousetka A, Lehrtrach H, Radelof U;
 XX MPI; 2000-195572/17.
 XX P-PSDB: AAY51611.
 XX
 XX Novel DNA encoding human HSGN1 useful for diagnosis or therapy of
 XX neurodegenerative disease, especially Smith Magenis Syndrome -
 XX
 XX Claim 1; Page 27-34; 39pp; German.
 XX
 XX This invention describes a novel human protein and its encoding nucleic
 XX acid, designated HSGN1 which has neuroprotective activity. The HSGN1
 XX protein, or antibodies raised against it are useful as reagents to
 XX diagnose and/or for therapy of neurodegenerative disease, in particular
 XX Smith Magenis Syndrome which is a multiple congenital anomaly. This
 XX sequence encodes the human HSGN1 protein described in the invention.
 XX
 XX Sequence 5395 BP; 1117 A; 1877 C; 1595 G; 806 T; 0 other;

alignment_scores:

Quality: 8479.00 Length: 1758
 Ratio: 5.189 Gaps: 6
 Percent Similarity: 92.947 Percent Identity: 92.776

alignment_block:

US-09-508-821b-6 x AA288975

Align seg 1/1 to: AA288975 from: 1 to: 5395

1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnGlnAsn 17
 1 ATGCAAGTCTTTCAGAAAGGTGGTTCCATGCGAAACAACAGAACTA 50
 17 GlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnPro 34
 51 CGAGCAGACCTCGCAGGAACATCAGGCTAGAAATTCAGAGCAGCCGA 100
 34 ArgGlnAsnGlyLeuSerCysAspArgGlnArgLeuAsnAlaLysAspTyr 50
 101 GTCAGGCGCGGCTAAGCTGCGACGCGGCTGCTCCGCCAAGACTAT 150
 51 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyThrProSe 67
 151 TATTAACCGCGACCTTACCGAGCTATAGAGGTGCGCTGGAGGCCCTC 200
 67 GlnTyrAlaAlaAlaAlaValAlaAlaAspLysTyrHisArgGlySerLys 84
 201 TGGCACTGACGAGCGCGGTGGCGCGCAGCAAGTACACAGCAGCAGCA 250
 84 LeuLeuProThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 100
 251 CCTGCGCCACAGACAGAGGCTGCGAGGGAGGCGGCTTCCCTGGCTAC 300
 101 GlnValGlnAspSerSerProTyrProGlyTyrTyrAlaGlyGlnGln 117
 301 GCGGTCCAGAGAGAGAGCCCTTACCCAGGCGCGCTATGCTGGTGGAG 350
 117 HisGlnAlaThrGlnAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 134
 351 CTTTCAGGCTTGGGGGCGCCCAAGCAGCAGCCGCCAGCGCGAGCCAC 400

134 eupProAlaGlyValAlaLysTyrAspGlnAsnLeuMetLysTyrThrAla 150
 401 TACCTGAGAGGGGTGGCCAAATGATGATGAACTGATGAAATAAGACACA 450
 151 ValProSerArgGlnTyrAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 167
 451 GGGCCCCCAGCAGAGAGTATGACAGAGGCGCGCCAGGTGCTCCCTTCG 500
 167 ThrHisSerLeuHisValGlnGlnProProProGlnGlnProLeuAla 184
 501 GACTCAGCTCCCTGACAGCTCCAGAGCCAGCCGCCGCCAGACCCCTGG 550
 184 LeuTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerPro 200
 551 CATACCCCAAGCTCCAAAGCAGAGAGCTGACAGAGACATTCGCTCCCT 600
 201 LeuProPheProGlnGlnTyrHisPheProGlnHisSerGlnSerPhe 217
 601 CTGCCCCCTTCCC..... 612
 217 ThrSerSerThrTyrSerSerSerValGlnGlyGlyGlnGlyAlaH 234
 613 634
 234 SerTyrLysSerCysPheAlaProThrAlaGlnProHisAspArgPro 250
 635 ACTCTATAGAGTTCACAGACAGAGCTGCTCCAGCCCAAGTACAGAGCG 684
 251 LeuThrAlaSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuH 267
 685 CTGACTGCGAGCTCCAGCTCGGCCCGCGGCGAGCGGTCCAGATCTTCA 734
 267 SalAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGln 284
 735 TCCCTACCGAGTGGGCGGCTGACCTATGACAGCAGCAGCAGCAGCAGC 784
 284 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
 785 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 834
 301 GlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlnGln 317
 835 GAAACCCCTCATTAACAAACCTCGCCCAATTCAGCAGCAGGAGAGA 884
 317 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 334
 885 AGCCAGGAGCTACTGCCAGCCGAGCAGGCGCTCCGAGCCAGACAGCAG 934
 334 TyrTyrGlnThrPheSerProSerSerSerHisSerProAlaArgSerVal 350
 935 ACTACAGACCTTACGCCAGCTCCAGCAGCTCCAGCAGCCGCCCTCCGTG 984
 351 GlyArgSerProSerTyrSerSerThrProSerProLeuMetProAsnLe 367
 985 GCGCGCTACCTTCCACAGTTCCACAGCTCCGCGCTGAGTACCAACT 1034
 367 GlnAsnPheProTyrSerGlnGlnProLeuSerThrGlyAlaPhePro 384
 1035 GAGAGACTTTCCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1084
 384 IsoleuIleThrAspHisSerHisPheMetProLeuLeuAsnProSerPro 400
 1085 CAGGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1134
 401 ThrAspAlaThrSerSerValAspThrGlnAlaGlnGlnGlnGlnGln 417
 1135 ACGGATGCCACAGCTCTCCT..... 1155
 417 GlnGlnAspLysLeuProGlnAsnLeuSerAspLeuSerLeuGlnHis 434
 1156 GCGAG 1159

Wed Dec 5 08:58:51 2001

434 erLeuThrAlaLeuThrLeuGlnValGluAsnHisSerAsnThrValGln 450
:::|||||
1160 ACCTCAGCGCGCTGACCTCAGAGTGGAGAACATCTCCACACCGTCCAG 1209
451 GlnLeuLeuLeuSerLysAlaAlaValProGlnLysLysGlyValLysAs 467
|||||
1210 CAGCTGCTGCTCTCCAGAGCTGCTGCTGCGCGCAGAAAGGTGTCAAGA 1259
467 nLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerProG 484
1260 CTTCTGTGTCAGAGCCCGCAGAGCAGCATAAAGCCAGCACTGCAGCCCG 1309
484 luGlySerGlyThrSerAlaGluProAlaGlyThrProLeuSerGluPro 500
1310 AAGGAGCGGCTACTCAGCGCGAGCCCGCAGAGCAGCAGCGTGCAGAGCGG 1359
501 ProSerSerThrProGlnSerThrHisAlaGluProGlnGluAlaAspTy 517
1360 CCGAGCAGCAGCCACAGTCCAGCATGCGGAGCGCGCAGAGCGCGACTA 1409
517 rLeuSerGlySerGluAspProLeuGluArgSerPheLeuThrCysAsnG 534
1410 CTTGAGCGGCTCCGAGGACCACTGGAGCGCAGCTTCTCTACTGCAACC 1459
534 lnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLysPro 550
1460 AGGCCGCTGGCAGCCCTGCCAGGCTCAACAGCACTCGAGGCGCAGAGCG 1509
551 GluSerValSerThrCysSerValThrSerProAspAspMetSerThrIly 567
1510 GAGTCCGCTGTCACCTGTTCTGTGACCTCTCTGAGCAGCATGTCCACCAA 1559
567 sSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAspSerP 584
1560 ATCTGAGCACTCTCCAGAGCCTACAGGCACTGTGCGCTCGACAGCT 1609
584 heSerLysPheValAlaGlyGluArgAspCysProArgLeuLeuLeuSer 600
1610 TCTCCAGATTCTGGCGGCTGAGCGGACGTCGCGGCTGCTGCTCAGC 1659
601 AlaLeuAlaGlnGluAspLeuAlaSerGluLeuLeuGlyLeuGlnGluAl 617
1660 GCTCTGGCAGAGAGAGCTTGGCTCCGAGATCTTGGGCTGCAGGAAGC 1709
617 alLeuGlyLysAlaAspLysAlaThrPallagluAlaProSerLeuVal 634
1710 CATCGTGAAGAGCGCAGAAAGCTTGGCTGAAGCAGCCAGCTGTCTCA 1759
634 yaAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysLeu 650
1760 AGGACAGCAGCAGCCACCTCTCTGCTGGAGAACCCAGCGCTGCTG 1809
651 AspSerValAlaLysSerAlaThrProArgProGlyGluProGluAlaLe 667
1810 GATCTGTGGCCAGAGTGGTGGCGCGCGCTGGGCTGGGAGCCGAGAGCC 1859
667 uProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerP 684
1860 GCCCGACTCTCTGAGCGTGCAGAAAGGCGGAATGCCAAGAGACTTCAGCC 1909
684 roGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProLys 700
1910 CAGGCTGTGTGAGAGCCCTCGGCTGGCTGCTGCTGCTGCTGCTGCTGCT 1959
701 LysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyValPr 717
1960 AAGACACTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2009
717 oAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrThr 734
2010 TGTCTCAGACCCCACTACAGCAGCTTTTGTACTGTCTCTCTCTCTCTCTCT 2059
734 laAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGluGluAsn 750

2060 CTGCCAGCTCAGCGGACAGCGCCACACCCCTTTGCTGGCCAGAGAAAC 2109
751 LeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuThrLy 767
2110 CTGGGGGATGCTTCTCCAGGTGGGATTCACCCCTGGCGAGCTTACCAA 2159
767 sGlyLeuGlnGlnGlyLysAlaSerAspGlyLysLeuLysGlyAspT 784
2160 GGGCTGGAGCAGGCTGGGAAGCCCTCAGATGGCATCAGCAAGGGGACA 2209
784 hrHisGlnAlaSerAlaCysLeuGlyPheGlnGlnGluAspProProGly 800
2210 CCTATGAGCTTCGGCCCTGCTGCTGGGCTTCAGAGGAGGACCCCTGGG 2259
801 GluLysValAlaSerLeuProGlyAspPheLysGlnGlnGluValGlyG 817
2260 GAGAAGGTGGCTCTGCTGCCCGGGACTTCAAGCAGGAGGAGGTGGTGG 2309
817 yValLysGluGluAlaGlyGlyLeuLeuGlnCysProGlnValAlaLysA 834
2310 GGTGAAGGAGGAGGAGGTGGGCTGCTGAGTGGCCCGGAGGTGCCCAAG 2359
834 laAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAspPhe 850
2360 CTGACCGGTGGCTGGAGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTG 2409
851 GlyAspLeuProLeuLeuProThrSerArgLysGluAspLeuGluAl 867
2410 GGGGACCTCCACTGCTGCCACCCAGCAGGAGGAGGAGCTTGAAGC 2459
867 aGluGluGluLysSerSerLeuGlyLeuLeuLeuGlySerProGluGlnA 884
2460 TGAGGAGGAGTACTCTCTCTATGTGAGCTCTCTGGGAGCGCGGCTGCC 2509
884 rgProGlyMetGlnAspProLeuSerProLysAlaProLeuLeuCysThr 900
2510 GGCTGGGATGCAGGAGCGCGCTGCTCACCAGGCGCGCTCATCTGCAAC 2559
901 LysGluGluValGluGlnValLeuAspSerLysAlaGlyTrpGlySerPr 917
2560 AAGGAGGAGTGGAGAGGTGCTGAGCTTCAAGCGCGGCTGGGCTCTCC 2609
917 cCysHisLeuSerGlyGluSerValIleLeuLeuGlyProThrValGlyT 934
2610 GTGCCACCTCTCAGGGAGTCCCTCATCTCTGCTGGGCGCTACAGTGGCA 2659
934 hrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMetLys 950
2660 CCGAGTCAAGGTCCAGAGCTGTTTGTGCTCTCTCTCTCTCTCTCTCT 2709
951 ProGlyGluGlyProAspGlyGluArgAlaProGlyAspSerThrTh 967
2710 CCAGGTCAAGAGGCGCTGATGGGAGCGAGCTCCAGGGATTCACACAC 2759
967 rSerAspAlaSerLeuAlaGlnLysProAsnLysProAlaValProGluA 984
2760 CTCGGAGCGCTCTCTGGCCCGAGAGCCACACAGCTCTGCTGCTGCTG 2809
984 laProIleAlaLysLysGluProValProArgLysSerLeuArgSer 1000
2810 GCGCCATCGCAAGAAAGAGCGCTGTCAGCGGGGCAAAAGCTTACGAGC 2859
1001 ArgArgValHisArgGlyLeuProGluAlaGluAspSerProCysArgAl 1017
2860 CGTGCGGTGACCGGGGCTGCCGAGGCGAGGACTCCCTCTCTCTCTCTCT 2909
1017 aProValLeuProLysAspLeuLeuProGluSerCysThrGlyProP 1034
2910 ACCAGTGTGCCCAAGAGCTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCT 2959
1034 roGlnGlyGlnMetGluGlyAlaGlyAlaProGlyArgGlyAlaSerGlu 1050

2960 CCCAGGAGCAGATGAAAGGGGCTGAGAGCCAGCGGAGGGGCTCCGAA 3009
1051 G1yleuProArgwecysThrArgSerleuThAlaLeuSerGluProAr 1067
3010 GGGCTCCAGAGATGTACTGTTCTCTCAGCGCCCTGAGAGAGCCCG 3059
1067 gthtProGlyProGlyLeuThrThThProAlaProPheAspLysL 1084
3060 CAGCGCCGAGACCCAGGCTGACACACCCCTGCAACCCCGAGCAAAAC 3109
1084 enuGlyGlyLeuGlnArgAlaAlaPheLysSerGlyLysArgValGlyLys 1100
3110 TGGGGGGCAGAGAGCGAGCGCTTCAAGTGGGGCAAGGGGTGGGAG 3159
1101 ProSerProLysAlaAlaSerSerProSerAsnProAlaAlaLeuProVa 1117
3160 CCTCACCCAAAGGCTGCTCCAGCCCGAGCAACCCGCGCCCTCCGT 3209
1117 AlaSerAspSerSerProMetGlySerLysThrLysGluThrAspSerP 1134
3210 GGCTCCGAGACAGACCCGATGGGCTCCAGACCAAGAGAGAGAGCTCAC 3259
1134 roSerThrProGlyLysAspGlnArgSerMetLLeuArgSerArgThr 1150
3260 CCAGCAGCGCTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3309
1151 LysThrGlnGluLeuPheHisSerLysArgArgArgProSerGluGlyAr 1167
3310 AAACCCAGAGAGATCTTCCACTCCAGCGCGAGCGCCCTGAGGGCG 3359
1167 GLeuProAsnGysArgAlaThrLysLysLeuAspAsnSerHisLeuP 1184
3360 GCTCCCACTGCGTGGCCAGCAAGAGAGCTCTCCAGCAACAGCCACTTC 3409
1184 roAlaThrPheLysValSerSerSerProGlnLysGluGlyArgValSer 1200
3410 CCGCCACATTCAGAGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 3459
1201 GlnArgAlaArgValProLysProGlyAlaGlySerLysLeuSerAspAr 1217
3460 CAGCGGGGAGGGTCCCAACCTGGTGCAGGAGAGAGAGAGAGAGAG 3509
1217 gProLeuHisAlaLeuLysArgLysSerAlaPheMetLapProValPro 1234
3510 GCCCTCCATGCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3559
1234 htlLysLysArgAsnLeuValLeuArgHisLysSerSerSerSerAsn 1250
3560 CCAAG 3609
1251 AlaSerAlaMetGlyGluMetGlyArgArgArgGlyLeuArgValProp 1267
3610 GCCAGTGGCAATGGGGAGATGGAGAGAGAGAGAGAGAGAGAGAGAG 3659
1267 roProSerSerArgGlyCysLeuLeuProArgLysProSerProArg 1283
3660 CACCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3709
1284 AlaMetAlaSerLeuProGlnSerSerHisProArgArgProPheL 1300
3710 GCATGGGAGAGCTGCCAAGAGCTCCAGCCCGAGAGAGAGAGAGAG 3759
1300 euProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLy 1316
3760 TGGCTCAAGCTGCGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 3809
1316 sValLeuProProArgLysGlyArgGlyLysLysLeuGlnAlaLeuValG 1333
3810 GGTGTGCGCAGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3859
1333 LysLysLysSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349
3860 AGAAGATCACCTCGCCAGAGCTCCAGAGAGAGAGAGAGAGAGAGAG 3909
1350 AlaSerProGlyAsnProLeuSerProSerLeuSerAspLysArgL 1366
3910 GCTCTCTCTGTAACTCTGAGCCCTCTTCCAGCAAGAGAGAGAGAG 3959
1366 yLeuLysGlyAlaGlyLysSerProValGlyValGluGlyLeuValA 1383
3960 GCTAAGAGTCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4009
1383 snValGlyThrGlyLysLeuProThrSerGlyAlaAspProLeuLys 1399
4010 ATGGGGCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4059
1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetLysSerLys 1416
4060 AGAATCCAAACACAGATCTTAAAGAGAGAGAGAGAGAGAGAGAGAG 4101
1416 sLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerProGlu 1433
4101 4101
1433 LeuGlnProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449
4101 4101
1450 GlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgPr 1466
4101 4101
1466 ctyLeuLysProAlaLeuLeuThrProArgAspArgAlaSerGlyT 1483
4102 AGGCGAGTGGCA 4114
1483 hrlGlnGlyAlaSerGluAspAsnSerGlyGlyGlyGlyLysLysProLys 1499
4115 CAAAGAGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4164
1500 MetGlnGluLeuGlyProAlaSerGlnProProGlnLysArgProCysG 1516
4165 ATGAGAGAGCTGGGCTGGGCTCCAGCGCCCGAGAGAGAGAGAGAG 4214
1516 hProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnLysSerS 1533
4215 GCCCGAG 4264
1533 ySerLysArgLysArgLeuThrArgGlyArgAlaLysAsnThrThrSer 1549
4265 ATTCAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4314
1550 SerProCysLysGlyArgAlaLysArgArgArgGlnGlnGlnValLeuPr 1566
4315 TCACCCCTGAAGGGCGTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 4364
1566 cLeuAspProAlaGluProGlnLysLeuArgLysLysLysSerSerCyl 1583
4365 CCGTGAATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4414
1583 ySArgLeuArgSerAspSerArgThrProAlaPheSerProPheValArg 1599
4415 AGCGGTGAG 4464
1600 ValGluLysArgAspAlaPheThrThrLecyThrValValAsnSerPr 1616
4465 GNGAG 4514
1616 oGlyAspAlaProLysProHisArgLysProSerSerSerAlaSerSers 1633
4515 TGAAGATGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4564
1633 eSerSerSerSerSerPheSerLeuAspAlaAlaCysLysAlaSerLeuAla 1649
4565 CCGTCT 4614


```

Percent Similarity: 100.000 Percent Identity: 99.351
Lignment block:
US-09-508-821B-6 x AAD05635
Align seg 1/1 to: AAD05635 from: 1 to: 2409
1602 LysArgAspAlaPheThrThrLeuValValAsnSerProGlyAs 1618
10 CGCGGAGAGCGTTTCCACCATATGCACTGTTGTCAACTCCCTGGAGA 59
1618 pAlaProLysProHisArgLysProSerSerSerAlaSerSerSers 1635
60 TGGGCCCAAGCCACAGAGAGCCCTCTCTGCTCTCTCTCTCTCTCAT 109
1635 erSerSerSerPheSerLeuAspAlaAlaLaglyAlaSerLeuAlaThrLeu 1651
110 CTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 159
1652 ProGlyGlySerIleLeuGlnProArgProSerLeuProLeuSerSerTh 1668
160 CTTGGAGGCTCATCTGCAGCGGGCCCTCTCTCTCTCTCTCTCTCT 209
1668 rMetHisLeuGlyProValValSerLysAlaLeuSerThrSerCysLeuV 1685
210 GATGCATCTGGGGCTCTGTGTTTCCAAAGCCCTGAGTACCTCTGCTTG 259
1685 alCysCysLeuGlyGlnAsnProAlaAsnPheLysAspLeuGlyAspLeu 1701
260 TTTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309
1702 CysGlyProGlyTyrrProGluHisCysLeuProLysLysLysLysLe 1718
310 TGTGGGCTCTACTACCTGACACTGCTCCCTCCCAAGAGCAAACT 359
1718 uLysGluLysValArgProGluGlyThrCysGluGluAlaSerLeuProL 1735
360 CAAGGAGAGGTTGGGGCCAGAGGACCTCTGTGAGAGGCTGCTGCTGC 409
1735 euGluAlaGlyThrLeuLysGlyProGluCysAlaAlaAlaThrAlaGly 1751
410 TTGAGAGACACTCAAGGTCCTGAGGTGCTGAGTGTGCTGCTGCTGCT 459
1752 LysProProArg 1755
460 AAGCCCTCCAGG 471
seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT: AAD05594
seq_documentation_block:
ID AAD05594 standard: cDNA; 766 BP.
AC AAD05594;
AC AAD05594;
DT 17-JUL-2001 (first entry)
DE Human secreted protein-encoding gene 16 cDNA clone HDPXY01, SEQ ID NO:26.
DE Human; secreted protein; proliferative disorder; cancer; tumour;
DE foetal abnormality; developmental abnormality; haematopoietic disorder;
DE immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
DE inflammation; allergy; neurological disorder; Alzheimer's disease;
DE Parkinson's disease; cognitive disorder; schizophrenia; asthma;
DE skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
DE cardiovascular disorder; angiogenic disorder; kidney disorder;
DE gastrointestinal disorder; pregnancy-related disorder; gene therapy;
DE endocrine disorder; infection; wound healing; vulnary;
DE cell culture; chemotaxis; food additive;
DE binding partner identification; ss.
XX Homo sapiens.
XX OS
XX FH
XX Key
XX Location/Qualifiers

```

1603 ArgAspAlaPheThrIleCysThrValValAsnSerProGlyAspAl 1619
 3 CGAGAGCGCTTCACACCAATATGCACTGTTGTCACTCCCTGGAAATGC 52
 1619 AProlysProHisArgLysProSerSerSerAlaSerSerSerSers 1636
 53 GCCCAAGCCGCCACAGAGAGCCTTCCTCTGCTCCCTCTCTCAATCC 102
 1636 eSerSerPheSerLeuAspAlaIaIaValaSerLeuAlaThrLeuPro 1652
 103 GCT 152
 1653 GlyGlySerIleLeuGlnProArgProSerLeuProLeuSerSerThirne 1669
 153 GGAGGCTCATCTCCAGACCGCGGCGCTCTCTCTCTCTCTCTCTCAAGAT 202
 1669 ThIleuGlnProValValSerLysAlaLeuSerThirneCysLeuValC 1686
 203 GCACTGGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 252
 1686 yScYsLeuCySgLnAsnProAlaAsnPhelYsAspLeuGlyAspLeuCyS 1702
 253 GCTGCTCTCTGCAAAACCGCGCAACTTCAAGAGACTTGGGGAACCTCTGT 302
 1703 GlyProTyrTyrProGlnHisCysLeuProLysLysLysProLysLeu 1719
 303 GGCGCT 352
 1719 sGlnLysValArgProGlnGlyThrCysGlnGlnAlaSerLeuProLeng 1736
 353 GGAGAGGCGCGCGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 402
 1736 LurGThrLeuLysGlyProGlnCysAlaAlaAlaAlaAlaAlaAlaAla 1752
 403 AGAGACACTCAAGAGGCTCGAGTGTGACCTCTCTCTCTCTCTCTCTCT 452
 1753 ProProArg 1755
 453 CCCCCCAGG 461

seq.name: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.AA073500

seq_documentation_block:
 ID AA073500 standard; DNA; 8438 BP.

XX AC AA073500;
 XX 15-MAR-1995 (first entry)
 DE DNA encoding Pseudorabies virus large latency transcript.
 XX Pseudorabies virus; PRV; LTV; large latency transcript;
 KW attenuated virus; vaccine; early protein 0; BP0; HSV-1 ICPO;
 KM protecting animals; deletion mutants; swine; ds.
 XX Pseudorabies virus.
 OS XX
 FH Key Location/Qualifiers
 FT misc_feature 1..7013
 FT /tag= a
 FT /note= "derived from PRV strain Inph"
 FT misc_feature 7014..8425
 FT /tag= b
 FT /note= "derived from PRV strain Ka"
 FT CDS 622..6498
 FT /tag= c
 FT /note= "encodes predicted amino acid sequence of ORF2"
 FT 1..6
 FT /tag= d
 FT misc_feature 34
 FT /tag= e
 FT /note= "RNA cap site"

FT polyA_signal 8382..8387
 FT /tag= f
 XX US3352596-A
 PN 04-OCT-1994
 PD 04-OCT-1994
 PF 11-SEP-1992; 92US-0945283.
 XX 11-SEP-1992; 92US-0945283.
 PR 11-SEP-1992; 92US-0945283.
 XX (USDA) US SEC OF AGRIC.
 PA Cheung AK, Wesley RD;
 XX PI
 XX DR MPI, 1994-316187/39.
 DR P-PDB: AAR60620.
 XX
 PT New pseudorabies virus mutants for use in vaccine - having a
 PT deletion and/or insertion in the early protein 0 gene or large
 PT latency transcript gene
 PS Disclosure: Column 15-30; 43pp; English.

XX AA073500 shows the pseudorabies virus (PRV) large latency transcript
 CC (LTV). The basic sequence is derived from PRV strain Inph and PRV
 CC strain Ka. The LTV overlaps and is transcribed in the opposite
 CC orientation with respect to the BP0 (early polypeptide 0) and the
 CC immediately early gene (IE180). BP0 is nonessential for replicatio,
 CC LTV is the only gene expressed during PRV latency, and the IE180
 CC gene is absolutely necessary for PRV replication. However there are
 CC 2 copies of IE180 in the genome. It is expected that PRV lacking one
 CC of the IE180 copies is viable. Deletions in the non-overlapping
 CC regions of these 3 genes will generate single deletion mutants,
 CC while deletions in overlapping regions will generate double deletion
 CC mutants. The invention is concerned with the construction of attenuated
 CC viruses which have a reduced ability to reactivate from latency. This
 CC can be achieved by functionally disabling the expression of the BP0
 CC gene, or by disrupting the synthesis of the LTV, or both. (See also
 CC AA073501 and AAR60620-24)

SQ Sequence 8438 BP; 1141 A; 2916 C; 3327 G; 1054 T; 0 other;

alignment_scores:
 Quality: 508.00 Length: 2060
 Ratio: 0.647 Gaps: 104
 Percent Similarity: 38.107 Percent Identity: 22.961

Alignment_block:
 US-09-508-821b-6 x AA073500/rev ..

Align seg 1/1 to reverse of: AA073500 from: 1 to: 8438
 31 ArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAl 47
 7236 CGGCGCGCATCTTGGCGCGGCGAGCA.....AAGCGAGCGAGANGA 7193
 47 alyAspTyrTyrAsnProGlnProTyrProSerTyrGlnGly..... 61
 7192 AGCGGCGCAGAAAAATCCATTTGGCGCGCGCGGAGAGTCCGCGGCA 7143
 62GlyAlaGlyThrProSerGlyThr 69
 7142 AATCGGCGCATTTGGCGCTTACCTGAGGCGGCGGCTCTCTCTCTCTCT 7093
 70 AlaAlaAlaValAlaAlaAspLysTyrHisArgGly..... 81
 7092SerLysAlaLeuProThrGln.... 88
 82
 7066 TTCACTGGCGGCGAGTAGAGACAGACATGTTCTCTGCAACCGAGGGG 7017

88 88
7016 ATCCGATCGTCTCCGCTCCGCGCGGACTCTGAAGACTCCGGCTCCGG 6967
89GlnGlyLeuGlnArgProAlaPheProGly.....Ty 100
6966 CGGATACAGCCCTCGAGGAGCCGCGCCACCCAGAGGCTCTCGGCCAC 6917
100 rGlyValGlnAspSerSerProTyrProGlyArgTyr..... 112
6916 CAGAGAGAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6867
113AlaGly.GluGlu... 116
6866 GCCCGCGCTCTCCGCGCTCCGCGCGCGCGCTCTCTCTCTCTCTCTCT 6817
116 116
6816 CCGATCGAGACCATGGCGGAGACTCTTTGACTTCATCCGAGCCGAGGG 6767
117SerLeuGlnAlaIrpGlyAlaProGlnProProProProGlnPro 131
6766 CAACCTCAGCAGCTCTGG.....CGCGCGCGCGCGCGCGCGCG 6726
132GlnProLeuProAlaGlyValAlaLysTyrAspGln 143
6725 AGGAGAGAGGATCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 6684
143 uAsnLeuMetLysLysLysLysLysLysLysLysLysLysLysLys 159
6683CTCTCCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCT 6641
159 LuGlnGlyAlaGlnValProPheArgThrHisSerLeuHisValGlnGln 175
6640 CCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6615
176 ProProProProGlnGlnProLeuAlaTyrProLysLeuGlnArgGln 192
6614 CCG 6579
192 sLeuGlnAsnAspIleAlaSerProLeuProPheProGlnGlyThrHis 209
6578CCGAGCCACGAGCGCG..... 6561
209 heProGlnHisSerGlnSerPheProThrSerSerThrTyrSerSer 225
6560CCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6534
226 ValGlnGlyGlyGlnGlyAlaHisSerTyr..... 236
6533 GGCACGCGCGCGCGCGCTCCGCGGAGATCGTGGTGTAGACGAGAGAC 6484
237 LysSerCysThrAlaProThrAlaGlnProHisAspArgPro..... 250
6483 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6434
251LeuThrAlaSerSerSerLeuAlaPro 259
6433 TCTCCATCAGGCTCCGAGCAGCGTCACTCTCTCTCTCTCTCTCTCT 6384
260 GlyClnArgValGlnAsnLeuHisAlaTyrGlnSerGlyValLeuSer 276
6383 GGGCGG.....GCTCGGGCGCGCGCGCG..... 6360
276 rAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAla 293
6360 6360
293 euGlnSerArgHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 309
6359 ..CGACCGCGGAGCACTCGCAGCG..... 6336

340 LysTyrGlnHisTyrGlyGlnGln..GlyGlnGlyTyrCysGlnProAsp 325
6335CAGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCC 6292
326 AlaAlaValArgThrProGlnGlnTyrTyrGlnThrPheSerProSer 342
6291 GCAGCC.....TCCCGCGCGCTC 6275
342 rSerHisSerProAlaArgSerValGlyArgSerProSerTyrSerSer 359
6274 CTTCT 6225
359 hrPro.....SerProLeuMet.....ProAsnLeuGluAsnPhe 370
6224 CCGCGCGAGATGTCCTCCCGGAGGGGAGCCCGTCTGAGCGCGCGG 6175
371 ProTyrSerGln.....GlnProLeuSerTh 379
6174 CCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6125
379 rGlyAlaPheProAla..... 384
6124 GCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6075
384 384
6074 CCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6025
385GlyLeuThrAspHisSerHisPheMetProLeu..... 395
6024 GACCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5975
396LeuAsnProSerProThrAspAlaThrSerSerV 407
5974 CGCGCGCGGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5925
407 alaThrGlnAlaGlyAsnCysLysProLeuGlnLysLysLysLeuPro 423
5924 TCTCT..... 5920
424 GluAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuThr 440
5919TCCACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5882
440 uGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeuSerLysA 457
5881 CGCGCTCGAGCC..... 5869
457 laAlaValProGlnLysLysGlyValLysAsnLeuValSerArgThrPro 473
5868GTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5854
474 GluGlnHisLysSerGlnHisCysSerProGluGlySerGlyTyrSerAl 490
5853 GAGAGAGAGAGAG.....GACGAGCTCGCGCTCCCGCGG 5822
490 aGluProAlaGlyThrProLeuSerGluProProSerSerThrProGln 507
5821 GGAC.....GGCGGGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 5782
507 erThrHisAlaGluProGlnGlnAlaAspTyrLeuSerGlySerGluAsp 523
5781CGCGCGCGGAGAGGGC 5767
524 ProLeuGlnArgSerPheLeuTyrCysAsnGlnAlaArgGly..... 537
5766 GCGCT 5717
538SerProAlaArgValAsnSerAsnSerLysAla..... 548
5716 CTCGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5667
549LysProGlnSerValSerThrCysSerVal 558

1046 rgGlyAlaSerGluGlyLeuProArgMetCysThrArgSerLeuThrAla 1062
 3996 CCGCGGCGCAAGCGCGCTCCCTGGAGTGGCGCGCTCCCGGACCCCG 3948
 1063 LeuSerGluProArgThrProGlyProGlyLeuThrThrProAl 1079
 3947 CTCGCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3913
 1079 aProAspLysLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1096
 3912 TCTCCGAGGACGACGCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 3875
 1096 ysArgValGlyProSerProLysAlaAlaSerSerProSerAsnPro 1112
 3874 CATGCCGAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3825
 1113 Ala.....AlaLeuProValAla 1118
 3824 CGGAGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3775
 1118 aserAspSerProMetGlySerLysThrLysGluThrAsp..... 1132
 3774 CCGCGGAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3725
 1133SerProSerThrProGlyLysAspGlnArgSerMet 1144
 3724 CTGGCG 3678
 1145 lleLeuArgSerArgThrLysThrGlnGluIlePheHisSerLysArg 1161
 3677CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3637
 1161 gArgProSerGluGlyArgLeuProAsnCysArgAlaThrLysLysLeu 1178
 3636 CGCGCTGATGCGGACATAGCCGCGCGCGCGCGCGCGCGCGCGCG 3587
 1178 euAspAsnSerHisLeuProAlaThrPheLysValSerSerProGln 1194
 3586 GCTCTAGACCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3537
 1195LysGluGlyArgValSerGlnAlaArgValProLysProGlyA 1210
 3536 CG 3490
 1210 laGlySerLysLeuSerAspArgProLeuHisAla.....Leu 1222
 3489 CTGCTCGCGCGCTTCGCCACCGCGCTCTGACGCGGAGCTGCACG 3440
 1223 LysArgLys..... 1225
 3439 GCGCGGGAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3390
 1226SerAlaPheMetAlaProValProThrLysLysArgAsn..... 1238
 3389 GGTGTGCTGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3340
 1239LeuValLeuArgHisGlySerSerSerSerSerSerSer 1252
 3339 TACCTGTGCTGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3290
 1253AlaMetGlyGluMetGlyArg..... 1259
 3289 CACATCGAGGACTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3240
 1260ArgArgGlyLeu. 1263
 3239 TCTACGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3190
 1264 ...ArgValPro...ProProSerSerArg..... 1271
 3189 CCGGAGTGGCGGAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3140

1272GlyCysLeuLeuProArg.....LysProSerProPhe 1282
 3139 GGGCGCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3090
 1282 oArgAlaMetAlaSerLeuPro.....Glns 1291
 3089 CCGGGCGCGAGCGCTGCGGCTCTGCCGCGAGGACAACTGGCTACACG 3040
 1291 erSerHisProArgArgProPro.....Phe 1299
 3039 GTGAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2990
 1300 LeuProGlnAlaArgLeuSerAlaAlaPheGlnGlyAlaMetLysThr 1316
 2989 CTACCGGAGCGCGCT..... 2975
 1316 sValLeuProProArgLysGlyArgGlyLeuLysLeuGluAlaIleVal 1333
 2974GCTCCCGACCGTGGACG.....CTGC 2953
 1333 InLysIleThrSerProSerLeuLysLysPheAlaCysLysAlaProGly 1349
 2952 AAGGACATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2906
 1350 AlaserProGlyAsnProLysSerProSerLysSerAspLysAspArg 1366
 2905 CTTTGACGCGGG.....CGCG 2889
 1366 yLeuLysGlyAlaGlySerProValGlyValGluGluGlyLeuVala 1383
 2888 CTTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2858
 1383 snValGlyThrGlyGlnLysLeuProThrSerGlyAlaAspProLeu 1399
 2857GCT 2855
 1400 ArgAsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLys 1416
 2854 CG 2828
 1416 sLeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProGlu 1433
 2827C 2827
 1433 laLeuInProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLys 1449
 2826 TCCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2777
 1450 GlyArgAla.....GlyAlaHisGlyLeuSerLys... 1459
 2776 GCGCGAGCTGCGCGCGCTCTGCCGCGCGCGCGCGCGCGCGCG 2727
 1460GlyProLeuLysArgProTyrLeuGlyP 1470
 2726 AGGCGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2696
 1470 roAlaLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAla 1486
 2695 CTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2646
 1487 SerGluAspAsnSerGlyGlyGlyLysLysProLysMetGluLeu 1503
 2645 TCGAGA..... 2640
 1503 uGlyProAlaSerGlnProGluGlyArgProCysGlnProGln...T 1519
 2639CCAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2594
 1519 hrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSer 1535
 2593 CTCG 2562
 1536 ArgLysArgLeuThrArgGlyArgAlaLysAsnThrThr..... 1548

```

2561 ..AGGACTGCGCGGCGGAGTCTGCGGCTCCGCGGCGGCGAGCGC 2515
1549 .....SerSerProCysLysGlyArgAlaLysArgArgA 1560
2514 CGGCCCCGCGGCGCGCTCGGCGCCATCAAGGAGGAGGAGCATCTCGAGCA 2455
1560 rGlnGlnGlnValLeuProLeuSppAlaGlnProGlnLileArgLeu 1576
2464 CGAGAGAG.....CCGAGGAGCGCGCGCAACCCCT..... 2436
1577 LysTyrLleSerSerCysLysArgLeuArg..... 1586
2435 .....ACCTGCTGCTCCGCTGAGCGGCGGCGCGCGCGCT 2395
1587 .....SerSerSerArgTyrThrProLalaphSerProPheValArgV 1600
2394 GGACTCTGACTGACTGACTCTCGGCGCGCGCGCGCGCTCGA. 2346
1600 aGlnLysArgAspAlaPheThrLleCysThrValValAsnSerPro 1616
2346 ..... 2346
1617 GYAAspAlaProLysProHisArgLysProSerSerSerAla...Serse 1632
2345 GGGCGCTTCTCT...CCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2239
1632 rSerSerSerSerSerSerPheSerLeuAspAlaAlaGlyAlaSerLeuA 1649
2298 GTCCCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2249
1649 lATHrLeuProGlyGly.....SerLleLeuGlnPro 1659
2248 CCTCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2159
1660 ArgProSerLeuPro 1664
2198 CGGCGCGGTGCGCT 2184
seq_name: /SIDS2/gcdata/geneseq/NA1997.DAT:AA789956
seq_documentation_block:
ID AA789956 Standard; DNA; 28958 BP.
XX
AC AA789956;
XX
DT 12-MAR-1998 (first entry)
XX
DE Sorangium cellulosum soraphen gene cluster genomic DNA.
XX
KW Soraphen; polyketide synthase; PKS; antipathogenic; antibiotic;
transgenic plant; phytopathogen; resistance; ss.
XX
OS Sorangium cellulosum.
XX
PN US562898-A.
XX
PD 02-SEP-1997.
XX
PF 01-JUN-1995; 95US-0457342.
XX
PR 20-AUG-1990; 90US-0570184.
PR 02-JUL-1992; 92US-0908284.
PR 31-AUG-1992; 92US-0937648.
PR 01-JUL-1993; 93US-0087636.
PR 08-JUN-1994; 94US-0258261.
XX
PA (CIBA ) CIBA GEIGY CORP.
XX
PI Beck JT, Gaffney TD, Hammer PE, Hill DS, Lam SF;
Ligon JM, Ryals JA, Schnupp T, Uknes SJ;
XX
DR WPI: 1997-447901/41.

```

```

XX
PR Protecting plants against pathogens with genetically transformed
PI biological control agent - which expresses all polypeptide(s)
PT involved in pyrrolnitrin biosynthetic pathway
XX
PS Ex 16; Column 104-128; 88pp; English.
XX

```

This genomic DNA sequence encodes the Soraphen gene cluster. This sequence encodes two open reading frames (ORF's), the positions of which are not given in the specification. ORF1 is approximately 25.5 kb in size and encodes 5 biosynthetic modules homologous to the erythromycin genes of *Saccharopolyspora erythraea*. Each module contains a beta-ketoadsintase (KS), an acyltransferase (AT), a ketoreductase (KR) and an acyl carrier protein (ACP) domain. ORF2 is immediately adjacent to ORF1 and is thought to contain 3 modules. Soraphen is a type I polyketide synthase (PKS) which has antibiotic activity. Transgenic plants containing such antipathogenic genes like those encoded in the Soraphen cluster should have enhanced resistance to attack by phytopathogens.

XX
SQ Sequence 28958 BP; 4084 A; 11863 C; 8356 G; 4655 T; 0 other;

alignment_scores:
Quality: 493.00 Length: 2095
Ratio: 0.593 Gaps: 102
Percent Similarity: 39.666 Percent Identity: 22.291

alignment_block:
US-09-508-821B-6 x AA789956

Align seg 1/1 to: AA789956 from: 1 to: 28958

```

62 GYAAlaGlyThrProSerGlyThrAlaAlaAlaAla.....Al 75
7509 GCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7658
75 aAspLysTyrHisArgLysSerLysAlaLeuProThrGln..... 89
7559 CCGTGGCGAGCATCGGCGAGCGAGCGAGTGAACAGAGAGAGAGAGAGAG 7708
90 ..GlyLeuGlnGly..ArgProAlaPheProGlyTyrTyrAlaGlnAspse 105
7709 GGGGCGGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 7758
105 rSer.....ProTyrProGlyTyrTyrAlaGlyGln 116
7759 CTCCAAAGCGGCGCTCTGCGGAGAGCGCCCTGCTGCTGCGGCGGCGG 7808
116 lSerLeuGlnAlaTyrPheLysAlaProGlnProProProGlnPro... 131
7809 GCTGCGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7858
132 .....GlnProLeu...ProAlaGln 137
7859 TCGACTGGGCGGCTTTCGCGCTTCCTTTCGCGGCGGCGGCGGCGGCGG 7906
137 YValAlaLysTyrAspGlnAsnLeuMetLysLysThrAlaValProProS 154
7907 ..... 7928
154 eArgGlnTyrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 170
7929 GCGGCGGCGGCGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7960
171 LeuHisValGlnGlnProProProProGlnGlnProLeuAlaTyrProLy 187
7961 .....CGTCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7984
187 sLeuGlnArgGlnLysLeuGlnAsnAspLleAlaSerProLeuProPheP 204
7984 ..... 7984

```


204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSerSer 220
 7985
 221 ThrTyrSerSerValGlnGlyGlyGlnGlyAlaHisSerTyrIy 237
 7997 ACAAGCTCGGAGCGCTCGGAGAGGAGGAGCTTCGTCGCTCGTCTCG 8046
 237 sserCys
 8047 CTGGTGGCCACAGACGCGCTCTCTCGGCGCAGGAGCGCTCCCA 8096
 247 isAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgVal 263
 8097 TGTCTGACCGGACAGAGGCTTCTCTGATC 8125
 264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280
 8125 8125
 280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 297
 8125 8125
 297 isHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313
 8126 8126
 314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328
 8144 TGGCGCTCGAGCTCGCGCGGCTTGCACAGAGGCGCCGCGCATCAAGCTC 8193
 328 ArgThrProGlnGln 340
 8194 CGGCGACCGCTCGCTTCGACCATCTCTCTCATCGAGTCTCGCTCTT 8243
 340 roSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
 8244 CTTCGCGAGTCTCGTCCGCCACGCGCTCGGAGGCTCTCGCTCGG... 8290
 357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...Proty 372
 8291 8291
 372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389
 8329 ACAGAGCGCCATCGCATCGTGGAGTGGCGCTCGCTCGCGGCGG 8378
 389 isSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404
 8379 CGTCGCGATGTCAGCGCTCTTGGAGTCTCTGCGCGAGGAGCGGAGC 8428
 405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420
 8429 GGGTGG 8429
 420 pLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThrA 437
 8455 TGGGATGCGCTGCTCTACAGCCCGGAGCGCGCGCGCGCGCGCGAG 8504
 437 laLeuThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeu 453
 8505 CTACGTCGCGGATCGCGCATCGTACAGGAGTGGAGTGGAGTGGAGT 8554
 454 LeuSerLysAlaAlaValProGlnLysLysGlyValLysAsnLeuValse 470
 8555 CCTCTTTGGCATCGAGCGCGCGGAGCGCGAGC...AC 8589
 470 ArgThrProGlnGlnHisLysSerGlnHisCysSerProGlnGlySerG 487
 8590 CTCAGCCCGC 8590
 487 LysSerAlaGlnProAlaGlyThrProLeuSerGlnProPro... 501

8619 ATCTGCTGGAGGCGCTCGAAGAGCGGCGCATCTGTCCTCCCGCCCTCA 8668
 502
 8669 AGATTTCCTCCCGCGCGCTCTGTCGCGCATCGCGCGCGCATAGCA 8718
 509 AlaGlu 516
 8719 TTGCGAGAGGCGGACACCGAAGATTCCGACGCTTATGCTCCCTCAAGGCA. 8767
 516 spTyrLeuSerGlySerGlnAspProLeuGlnArgSerPheLeuTyrCys 532
 8768 8803
 533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLys 549
 8804 CGTCTGCGCTCGAAGGCGCG...CGCTCGGTCGACACCGCGCTCTCC 8850
 549 sProGlnSerValSerThrCysSerValThrSerProAspAspMetSert 566
 8851 TCTCTGCTGCTGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 8900
 566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
 8901 GTGCAACCTCGCGCTCGCGCGGCGCTCTCGCTCATGCTCGCTCGCTCG 8950
 583 SerPheSerLysPheValAla 589
 8951 GCTTCGCT 9000
 590 595
 9001 AAGACCTTCTCGGCGCAACGCGCGCTCGGAGCGGAGCGGAGCGGAGC 9050
 595 roArgLeuLeuLeuSerAlaLeu 608
 9051 CCGCT 9100
 609 SerGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 625
 9101 TCTCTGCT 9150
 625 a 631
 9151 GGTATCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9200
 631 erLeuValLysAspSerSerLysProProPheSerLeu 643
 9201 GCTCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9250
 644 653
 9251 ATGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9289
 653 laLysSerAlaThrProArgProGlyGlnProGlnAlaLeuProAspS 670
 9290 9325
 670 erLeuGlnLeuAspLysGlyLysAlaLysAspPheSerProGlyLeu 686
 9326 9349
 687 PheGlnAspProSerValAlaPheAlaThrProAspProLysLysThr 703
 9350 CGCTCAAGACCAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9399
 703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720
 9400 GTGCGCAAGATCG 9443
 720 spProThrThrAlaAlaPheAspCysPhePro AspThr 732

9444 CACCTTCACACGAGGCGCCGCAAT...CCCTTGATGATTGGATACA 9489
733 ThrAlaIaSerSerAlaAspSerAlaAsnProPheAlaTrpProGlu 749
||| : : : : : ||| : : : : :
9490 CTGCGCATCGAGCTGTTGATACC...CCGAGTCTTGAGGCGCCCA 9533
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuT 766
: : : : :
9534 GAA...GATAGCATCCCGC... 9552
766 hrlYsGlyLeuGluGluGlyLysAlaSerAspGlyIleSerIysGly 782
: : : : :
9553 : : : : : CGCGCGCGCTCCGCTTCGACCTCCGACCTCCGCGCAC 9588
783 AspThrHis...GluAlaSerAlaCysLeuGlyPheGlnI 795
: : : : : ||| : : : : :
9589 AACGCCACGTCATCTCGAGAGAGGCTCCCGCGCCCTG...TC 9629
795 uGluAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLysG 812
: : : : : : : : : : :
9630 GAGCGAGCGCGCACCTCAGAGCGGCTCGGACCGCTCCCGCGCGCT 9679
812 lngGluValAlaGlyValLysGluAlaGlyLysLeuGlnCys 828
: : : : : : : : : : :
9680 GTCCCGTCTCTGTCGCGCAGAGCGCGCGCTCCCGCGCGCGAGCG 9729
829 ProGlu...ValAlaLysAlaAspArgTrpLeuGln 840
: : : : : ||| : : : : :
9730 AAGCGGCTCGCGCACCACTCTCCGCGACAGACCTCCGCTTATACA 9779
840 pSerArgHisCysCysSerThr...AlaAspPhe.GlyAspLeuPro 854
: : : : : : : : : : :
9780 TGTGGCTATTCGAGCGCACACCGCGCCACTTCGAGCACCGCGCG 9829
855 LeuLeuProProThrSerArgLysGluAspLeuGluAlaGluGlnI 871
: : : : : : : : : : :
9830 CTCTCTGCGCGCGAGCGCGAGC... 9853
871 rSerSerLeuCysGluLeuLeuGlySerProGluGlnArgProGlyMetG 888
: : : : : : : : : : :
9854 .AGCTCTCTCCGCGCTCGACTCGCTCGCCAGAGACACACCG... 9895
888 lAspProLeuSerProLysAlaProLeuIleCysThrLysGluGlnVal 904
: : : : : : : : : : :
9896 : : : : : : : : : : :
905 GluGluValLeuAspSer...LysAla...GlyTr 914
: : : : : : : : : : :
9932 GCAAGGTCTCTGCTCTTCCTTCGCGCAAGGCTCGCAGTGGAGAGATG 9981
914 pGlySerProCysHisLeuSerGlyGlnSerValIleLeuLeuGlyProT 931
: : : : : : : : : : :
9982 GCCCTCTCCCTGCTGACTCTCGCGGCTTCGCGCTCAGCTGAGAGC 10031
931 hrValGlyThrGlnSerIysValGlnSerTrpPheGlnUserSerLeuSer 947
: : : : : : : : : : :
10032 ATGCGAGCGCGCGCTCGCTCTCAGCTG... 10060
948 HisMetLysProGlyGluGluGlyProAsp...GlyGluArgAlaTr 962
: : : : : : : : : : :
10061 : : : : : AGTGAAGCTGCTCGCGCTCTCGCGCGAGAGGCGCGC 10101
962 oGlyAspSerThrThrSerAspAla... 970
10102 CCTCCCTCGACCGCGTGCAGCTGACAGCCGCCCTCTTGCGCTCAT 10151
971 : : : : : : : : : : :
10152 GATCTCCCTGCGCGCGCTCTGCGCTGCTCGAGCGCGCGCGCGCG 10201
982 ProGluAlaProIleAlaLysGluProValProArgLysSerIle 998
: : : : : : : : : : :
10202 TGTGTGGCCACAGCCAGGCGCGAGATCCCGCGCGCTCTGTCAGAGGCGCT 10251

998 uArgSerArgArgValHisArgGlyLeuProGluAlaGluAspSerProC 1015
: : : : : : : : : : :
10252 CTCTCCCTCGAGAGCGCGCGCGCATCCGCG...CCCT 10286
1015 ys...ArgAlaProValLeuProLysAspLeuLeuProGlu 1028
: : : : : : : : : : :
10287 GCGAGAAAGCGCTCACCGCGCGCGCGAGAGCGCGCGCGCGCGCG 10336
1029 SerCysThrGlyProProGluGlyLys...MetGlnIysAla.. 1041
: : : : : : : : : : :
10337 TCGAGCTCGCGCGCTCCGACCTCCAGACCTACCTCGCTCCCTGCGCGAGC 10386
1042 : : : : : GlyAlaProGlyArgGlyAlaSerGln 1051
10387 AGGCTCTCCACCGCGCGCGCTCAACCGCGCGCGCTACCTCGATCCG 10436
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerLuproArg 1067
: : : : : : : : : : :
10437 CGAGCGCGCGCGCGCTCGAGCGCGCTCTCGAGCTCTCACCGCGCACAAAG 10486
1068 : : : : : ThrProGlyProProGlyLeuThrThrProAlaPro... 1080
10487 TGTTCGCCCGCAAGATCCGCGTGCACCTACGCTCCACCTCCCGAGATG 10536
1081 : : : : : ProAspLysLeuGlyG 1086
10537 GAGCGCTCAAGACGAGCTCGCGCGAGCTACGACATGCTCTCG 10586
1086 LysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
: : : : : : : : : : :
10587 GAGCTCGAGCTCTCTCTTAT...CGACCGTCA 10618
1103 ProValAlaIaSer...SerProSerAsnProAla... 1113
10619 CCGGACACAGGCTCGAGCGCTCCGAGCTCGAGCGCGCTACGTATGTA 10668
1114 : : : : : AlaLeuProValAlaSerAspSerProMetGlySerLysT 1128
10669 AACCTCGGCAACCGCTCTCTGAGCGCGAGCGAGGCTCTCTGTA 10718
1128 hrlYsGluThrAspSerProSer... 1135
10719 CGAGGCGCATCGCTCTCTCGAGAGCTACACCGCCACCTCCGTCACGC 10768
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLysT 1152
: : : : : : : : : : :
10768 TCGCCCTCCCGAGACCTCGAGCGCTCACCGCTGATCCCGCTGCTGTC 10818
1152 lngGlnIlePheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168
: : : : : : : : : : :
10819 G...CTCATTCGAGAGAGAGAGCGCACTCG... 10849
1169 ProAsnGys...ArgAl 1173
10850 CCGCGCTGCTCTCTCTCTGAGAGCTCTACCGAGGCTCGCGCGCTC 10899
1173 ThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysAla 1190
10900 GACTGAGAGACTCTCTCGCGCGCTCGAGCTCCCGGAGAGTCTCTCTC 10949
1190 erSerSerProGlnLysGlu...GlyArgLysSerGlnAla 1203
: : : : : : : : : : :
10950 CACCTACCCCTTCAGAGAGAGAGGCTTGTGCTGAGCTCTCAGAGAGC 10999
1204 ArgVal... 1205
11000 AACGCTCCGAGCTCGCTCCGAGCGCTGACCTCGCGCGAGACATCCG 11049
1206 : : : : : ProLysProGlyAla... 1210
11050 CTGCTCGCGCGCGCGCTCGAGCTCGCGAGCGCGCTCTCTCTCTTAC 11099

1211GlySerLysLeuSerAspArg 1217
11100 AGCGGGCTCTCCCTCGAGAGCACCCTGGCTCGAAGGCCATCGCGTCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGCACACCATCTCTACCGGGCAGCGCTTCTCGAGCTCGCGCTGCAC 11199
1229MetAlaProValProThrLys.....LysArgA 1238
11200 GTCCGCCACCGCTCGCGCTCGACACCGCTCGAAGAGCTCAGCTCGAGGC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerSerSerSerSerSerSer 1254
11250 CCTCTCGCTCTCCATCGCAGGACACCTCTCTCTCTCTCTCTCTCTCTCTCT 11299
1255 Gly..... 1255
11300 GCGCGTGGAGCAGCAGGAGCAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 11349
1256GluMetG 1258
11350 GAGGACGGCTTCAGGATGGCCCTGGACTCGCCAGCCAGCGCTCTCT 11399
1258 LysArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274
11400 CTGGCGGGCAGCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11443
1275 LeuProArg.....LysProSe 1280
11444 CTCCCTCGAGTGCCATCCCGGTGACCTCGAAGGCTCTACGCAACCTC 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCAACCTCGGGCTTGCTACGCGCCGAGTTCAGGGCTCTCGCTCCGT 11543
1290 lnSerSer.....His 1293
11544 CTCACAGCGCGGACAGAGCTTTTCGCAAGCAGCAGCTCCCGGAGCGG 11593
1294 ProArgArg...ProProPheLeuProGlnAlaLeuArgLeuSerAlaAlaPhe 1309
11594 CCGAAAGGATCGCGCGGTTTGCC...TCCACCTCGCTGCTCGAC 11640
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326
11641 AGCGCCCTCTCATGCACTGSCCTTTGAGGAGCAGCAGAGGAGCGCTGC 11690
1326 eu..... 1326
11691 TCTGCCCTCTCTGGAGCGAGTCTCGCTCGCTCGCTCGCTCGCAACA 11740
1327LysLeuGluAlaLeuValGlnLysIleThrSerProSer..... 1339
11741 CTTTGGCGGTGCGCTTCCACCGTCCCAAGGTGAATCTCTCGCTCGATC 11790
1340LeuLysLysPheA 1344
11791 GTCTGGCCGACGCGCGAGTGAACCTCTTCTCTCTCTCTCTCTCTCTCTCT 11840
1344 laCys..... 1345
11841 CATGGGAGAGCTCGCGCGGAGCTCCGACCCCGGAGCTTCCACC 11890
1346LysAlaPro..... 1348
11891 ATGATGGCTCTTCCGCGTCTGAGTGGAGGAGCTCCAAAGCCCACTTCA 11940
1348 1348
11941 CGGCTCGCGCCCGGAGGCGCTCTCTCTCGGACAGGCGGCGCAGATCT 11990
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360

11991 CGCGCTCGAGCCCGCTCGCCGCTACGCGGACCTCGCTCCCTCCGAA 12040
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyLysProValG1 1376
12041 GCGCTCTGACAGGGGCTTGGCTCCCGGCTCGCTCGTGGCC..... 12085
1376 yValGluGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392
12086CCTTCATCATGATCGACCGGAGCGCTCTCTCTCTCTCTCTCTCTCTCT 12131
1393SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404
12132 GGCACCGCGCTCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421
12192 GCTCGCTCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12231
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGly 1438
12232 CACACCGAAGACGCTCAAGGACCTCG.....CTCAGCGCGCTCTCTG 12275
1408 LysThrAlaLeuAlaProLysArgSerArgLysGlyArgAlaGlyAla 1454
12276 GGGGCTCGCGCT 12325
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrlLeuGlyProAl 1471
12366 TCGTCCGACATCG.....ACCTCAGCGAGGCTCTCCAGCAGGCGCTG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 CTAGGCGCTCGACACAGGAGAACCCAGCTCTCGCTCTCGCAAGGAA 12416
1488 luAspAsnSerGlyGly..... 1493
12417 ACCCTCATCCCGAGTTGGGCGCAACACGCTCGAGGAGCGCTCTATCC 12466
1494GlyGlyLys.....LysProLysMetGluG1 1502
12487 GCGCGCAAGCAGGCTGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12516
1502 uLeu...GlyProAlaSerGlnProProGluGlyArgProCysGlnProG 1518
12517 GACGCTCGCT 12560
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrlSerSerTyrlSer 1534
12561 CGCACAGCGCAAGTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12610
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
12611 GCGATGCTCTGACACCTCTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12658
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561
12659TCGAGCGCAAGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12692
1561 lnGlnGlnValLeuProLeuAspPro..... 1569
12693 CGGTCTGAGTCT 12742
1570 ...AlaGluProGluIleArgLeuLysTyrlSerSerSerCysLysArgLe 1585
12743 TCGGCGGAGCTTTGTC..... 12760
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602
12761 .CCAGGCGCATCGCGAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 12790
1602 ysArgAspAlaPheThrThrIleCysThrValValAsnSerProGlyAsp 1618

7997 ACAAGCTCCGAGCCGCTCGAGAGGACGACCTTCCTGCTGCTCG 8046
237 sSerCys.....ThrAlaProThrAlaGlnProH 247
8047 CTGGTGGCCGACGAGCGCCCTCTCTCGCCACGAGCGCTCCCA 8096
247 sAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgVal 263
8097 TGTGACCCCGCAAGGCTTCCTCGATC..... 8125
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnI 280
8125 8125
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 297
8125 8125
297 sHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 313
8126TCGGTCTCGATTCGCTCA 8143
314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328
8144 TGGCCGTCGAGCTTCGCGCGCTTCGACACAGCGCCACCGCATCAAGCTC 8193
328 lArgThrProGlnGln.....TyrTyrGlnThrPheSerP 340
8194 CGGCGCCACCTCGCTTCGACCATCCCTCTCATGAGTCGCGCTT 8243
340 rSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
8244 CTTGCGGAGCTCGCTCGCGCACGCGCTTCGACACAGCGCTTCGCTG 8290
357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...ProTy 372
8291AGCCGACCGCGCGCTTCGCGCGCTTCGCGCGCG 8328
372 sSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389
8329 AGCAGCAGCGCATCGCATCGTCGATGCGCTTCGCGCGCTTCGCGCGCG 8378
389 sSerHisPheMetProLeuAsnProSer...ProThrAspAlaThr 404
8379 CGTGGGATGTCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCG 8428
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420
8429 GCGTCG.....AGCCGATTCGACGAGCGCG 8454
420 pLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThra 437
8455 TGGGATCGCGCTTCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCG 8504
437 lAlaLeuThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeuLeu 453
8505 CTACCTCGCGCAGCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCG 8554
454 LeuSerLysAlaAlaValProGlnLysGlyGlyValLysAsnLeuValSe 470
8555 CTTTGTGTCATGACGCGCGCGCGCTTCGCGCGCTTCGCGCGCG 8589
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGlyLysSerG 487
8590 CTCGAGCGCC.....AGCAGCGCTTCGCTTC 8618
487 lYThrSerAlaGlnProAlaGlyThrProLeuSerGlnProPro..... 501
8619 ATCTGCTGCGAGCGCTTCGAGACGCGCGCTTCGCGCGCTTCGCGCGCG 8668
502SerSerThrProGlnSerThrHi 509
8669 AGGATTCGCGCGCGCTTCGCTGCGCATCGCGCGCGCGCGCGCGCGCA 8718

509 sAlaGlu.....ProGlnGlnAlaAla 516
8719 TTCGAGAGGCGGACGACCGAAGATTCCGAGCTTATGCTCCGAGGCA 8767
516 sPyrLeuSerGlySerGluAspProLeuGlnArgSerPheLeuTyrCys 532
8768CGCGCGGCTTCCTTCGCGCGCGCGCTTCGCGCGCG 8803
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaL 549
8804 CGCTCGGCTTCGAGGCGCG.....CGCTTCGCTGCGACACCGCTGCTCC 8850
549 sProGlnSerValSerThrCysSerValThrSerProAspAspMetSer 566
8851 TCCTCGCTGCTGCGCTTCGACCTCGCTTCGCGCGCGCGCTTCGCGCGCG 8900
566 hLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
8901 GTGCAACCTCGCTTCGCGCGCGCGCTTCGCGCGCGCGCTTCGCGCGCG 8950
583 SerPheSerLysPheValAla..... 589
8951 GCTTCGCTTCCTTCGCGCGCGCGCTTCGCGCGCGCGCTTCGCGCGCG 3000
590GlyGluArgAspCysP 595
9001 AAGACCTTCGCGCGCGCGCGCTTCGCGCGCGCGCGCTTCGCGCGCG 9050
595 rArgLeuLeuLeuSerAlaLeu.....AlaGlnGlnAspLeuAla 608
9051 CGCTTCGCTTCGCGCGCGCGCTTCGCGCGCGCGCTTCGCGCGCGCG 9100
609 SerGluIleLeuGlyLeuGlnGlnAlaIleGlyGlnLysAlaAspLysAl 625
9101 TCCTCGCTTCGCTTCGCGCGCGCGCTTCGCGCGCGCGCTTCGCGCGCG 9150
625 a.....TrpAlaGlnAlaProS 631
9151 GGTATCAGCGCGCGCGCGCGCTTCGCGCGCGCGCGCTTCGCGCGCG 9200
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643
9201 GCTCCACGAGCGCGCGCTTCGCGCGCGCGCGCTTCGCGCGCGCG 9250
644GluAsnHisSerAlaCysLeuAspSerVa 653
9251 ATGCGACGCGCGCGCTTCGCGCGCGCGCGCTTCGCGCGCGCG 9289
653 lAlaLysSerAlaThrProArgProGlyGluProGlnAlaLeuProAspS 670
9289AAGCGCTTCGCGCGCGCGCTTCGCGCGCGCGCG 9325
670 erLeuGlnLeuAspLysGlyAsnAlaLysAspPheSerProGlyLeu 686
9326CTGAAAGCGCTTCCTTCGCGCGCG 9349
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThrTh 703
9350 CGCTCAAGACCAACATCGCGCGCGCGCTTCGCGCGCGCGCGCG 9399
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720
9400 GTGCGCAAGATCG.....TCGCGCTTCGCGCGCGCGCGCGCGCG 9443
720 sProThrThrAlaAlaPheAspCysPhePro.....AspThr 732
9444 CACCTCCACACGCGCGCGCGCGCGCAAT...CCCTGATTGATTGGGATACA 9489
733 ThrAlaAlaSerSerAlaAspSerAlaAsnPropheAlaTrpProGluGl 749
9490 CTCGCCATCGAGCTTCGTTGATACC.....CCGAGGCTTCGCGCGCGCGCA 9533

749 uasnleuGlYAspAlaCysProArgTrpAlaLeuHisProGlyLeuLeuT 766
766 hrlYsgLYleuGluGlnGlyLysAlaSerAspGlyIleSerLYsgLY 782
9553CGGCGCGGCTCTCGCCTTCGGACCTCTCCGACAC 9588
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnG 795
9589 AACGCCACACGATCCTCGAGAGAGGCTCCGCCGCCCTG.....TC 9629
795 uGluAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLYsg 812
9630 GGGCGAGCCCGCACCTCACAGAGGCGCTCGACCGCTCCCGCGCGCT 9679
812 InGluGluValGlyGlyValLYsgLysGlnGluGlyLYleuLeuGlnCys 828
9680 GTCCCGTGTCTCTGTGGCGAGAGAGCGCGCTCGCGCGCGAGGCG 9729
829 ProGlu.....ValAlaLYsAlaAspArgTrpLeuGlnAs 840
9730 AAGCGCTCCGCGACACCACTCCTCGCCAGAGACGACCTCGCTATCGA 9779
840 pSerArgHisCysCysSerTrp.....AlaAspPhe.GLYAspLeuPro 854
9780 TGTGGCTATTCGACAGCCACCAACCCCGCCACCTTCAGACCGCGCG 9829
855 LeuLeuProProThrSerArgLYsgLYAspLeuGluGlnGluGlnTyr 871
9830 CTCTCTGTGGCCGCGACCGCGACG..... 9853
871 rSerSerLeuGlyGlyLeuLeuGlySerProGluGlnArgProGlyMetC 888
9854 .AGCTCTCTCCGCGCTCGCTCGCTCGCCAGAGACAAAGCCG..... 9895
888 InAspProLeuSerProLYsAlaProLeuIleCysTrpLYsgLysGlnVal 904
9895CCCGAGACCGATTCGCGCGCGAGCGAGAAAGCGACG 9931
905 GlnGluValLeuAspSer.....LYsAla.....GlyTrp 914
9932 GCAAGGTGCTCTGCTCTCTCGTGGGCAAGGCTCGCAGTGGAGAGGATG 9981
914 pGlySerProCysHisLeuSerGlyLysValIleLeuLeuGlnLYsTrp 931
9982 GGCCTCTCCCTCGCTCGACCTCCGCGCGCTCTTCGCGCTCAGCTCGAAGC 10031
931 hTrpAlGlyThrGlnSerLYsValGlnSerTrpPheGlnSerSerLeuSer 947
10032 ATGCGAGCGCGCTCGCTCGCTCGACGTCG..... 10060
948 HisMetLYsProGlyGluGlnGlyProAsp.....GlyGluArgAlaPro 962
10061AGTGAACCTGCTCGCTCGCTCGCGCGCGAGAGGAGGCGC 10101
962 oGlyAspSerTrpThrSerAspAla..... 970
10102 CCTCTCTCGACGCGCTCGACGCTGTACAGCCGCCCTCTTGGCCGTAT 10151
971SerLeuAlaGlnLYsProAlaLYsProAlaVal 981
10152 GGTCTCCGCGCGCGCTCTGCGCTGCGCGCGAGCGCGCGCGCGCG 10201
982 ProGluAlaProIleAlaLYsGlyLysGlnProValProArgGlyLYsSerLe 998
10202 TCGTGGCGCACAGCCAGGCGAGATCGCGCGCTTCGTGCGAGCGGT 10251
998 uArgSerArgTrpValHisArgGlyLeuProGluGlnGlnAspSerProC 1015
10252 CTCTCCCTCGAGAGACCGCGCGCATCGCG.....CCTT 10286
1015 ys.....ArgAlaProValLeuProLYsAspLeuLeuProGlu 1028

10287 GCGAGAGAAAGCTCTACACACCTCGGCGGCAAGGCGGATGCGCGCG 10336
1029 SerCysThrGlyProProGlnGlyGln.....MetGlnLYsAla.. 1041
10337 TCGAGCGCGCGCTCGACCTCCAGACCTTACCTGCTCCCGGAGGAC 10386
1042GlyAlaProGlyArgGlyAlaSerGln 1051
10387 AGGCTCTCCACCGCGCGCTACAGCCCAAGGCGCTACCTGTATCGG 10436
1051 LYleuProArgMetCysTrpArgSerLeuThrAlaLeuSerGlnProArg 1067
10437 CGAGCGCGCGCGCTCGAGCGGCTGCTCGACGCTCTACCGCCGCAAG 10486
1068 ...ThrProGlyProProGlyLYleuThrTrpThrProAlaPro..... 1080
10487 TGTTCGCGCGAGATCCGCTGAGACTAGCGCTCCACTCGCGCCAGATG 10536
1061ProAspLYsLeuGlyG 1086
10537 GAAGCGCTCCAGAGAGCTCGCGCGAGGCTTACAGCAATCGCTCTCG 10586
1086 LYsGlyGlnArgAlaAlaPheLYsSerGlyLYsArgValGlyLYsProser 1102
10587 GACGTGGAGCTCTCCTTATT.....CGACCGTCA 10618
1103 ProLYsAlaAlaSer...SerProSerAspProAla..... 1113
10619 CCGGACACAGCTCGACGCTCCAGCTCGAGCGCGGCTACTGTATCGA 10668
1114AlaLeuProValAlaSerAspSerProMetGlySerLYsTr 1128
10669 AACCTCCGCGCAACGCTCTGTCTGTGAGCGGAGACCGAGGCGCTCGA 10718
1128 hTrpLYsgLYThrAspSerProser..... 1135
10719 CGATGGACATCGCTCTCGGCGAGGTACAGCCCGCATCGCTCGACG 10768
1136 ThrProGlyLYsAspGlnArgSerMetIleLeuArgSerArgTrpLYsTr 1152
10769 TCGCCCTCGCGAGACTCGAGGCGCTCACCGCTGATCCGCTGCTGTC 10818
1152 hGlnGlnIlePheHisSerLYsArgArgArgProSerGlnGlyArgLeu 1168
10819 GG.....CTCATTCGACGAGAGAGGCGACCTCG..... 10849
1169 ProAsnCys.....ArgAla 1173
10850 CCGGCTGCTCTCTCTCGGCGGAGCTCTTACCGAGGCTCGGCGCTC 10899
1173 ArgTrpLYsLeuLeuAspAspSerHisLeuProAlaThrPheLYsValS 1190
10900 GACTGGAGAGACTTCTTCGCGCTTACGCTCCCGCAGAGTCTCTCC 10949
1190 eSerSerProGlnLYsGln.....GlyArgValSerGlnArgAla 1203
10950 CACCTACCCCTTCAGAGAGAGCGGTTCTGGCTCGAGCTCGACAGAG 10999
1204 ArgVal..... 1205
11000 AACGCTTCGAGCTGCTCGACAGGCTTACGCTCGAGCGCCATCCG 11049
1206ProLYsProGlyAla..... 1210
11050 CTGTTCGCGCGCGCTGCTGCGCTTACGACCGCGCTTCTCTTAC 11099
1213GlySerLYsLeuSerAspArg 1217
11100 AGGCGGCGCTCTCCCTCGAGAGACACCGCTGCTCGAGAGCCATGCGCT 11149
1218 ProLeuHisAlaLeuLYsArgLYsSerAlaPhe..... 1228

[illegible]


```

1629 .....SerAlaSerSer ... 1633
12865 CTGTGCGATCTGGGGCATCTGGAACCCCAATAGCGTCTTAATCATTGC 12911
1634 .....SerSerSerSerSerSerLeuaspAlaAlaGly 1645
12915 GCGCGCGGGCGGCTCGGGACGGCGCGCTTGACTTCGACGCCACCTCG 12966
1646 AlaserLeuAlaThrLeuPro.....GlyGlySerLeuLeuGlnPr 1659
12965 GGCGCGAGGTCTTTGGCCACCGCCAATCCAGGAGAAAGTAGACCTCTCCGC 13011
1659 cArpProserLeuProLeuSerSerThrMetHisLeuGly..ProValy 1675
13015 GCGGCTGG.....GCTTCACGACANCGGACCTCGCTCCTCACG 13052
1675 alserLySAla...LeuSerThrSerCys 1683
13053 TGACGCGSGCTTCGACGAGACACTCTCTCTC 13081

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA1998.DMT.AAV05287

```

ID AAV05287 standard; DNA; 49377 BP

AC	AAV05287;
XX	
DT	21-MAY-1998 (first entry)

DE The soraphen biosynthesis gene cluster from *Sorangium cellulosum*.
XX
XX Polyketide synthase; PKS; biosynthetis; soraphen; SOR; SORa; SORb;
KM SORc; biosynthetis module; beta-ketosynthetase; acyltransferase;
KM ketoreductase; beta-ketone processing domain; cytosolic agent;
KM antimicrobial agent; phytopathogenic fungi; transgenic plant;
KM biological control; ss.
XX
XX *Sorangium cellulosum*.
DS

EH	Key	Location/Qualifiers
CDS		383..760
FT	/*cag- a	
FT	/product= Soir	
FT	/note= "gene product highly homologous to the	
FT	reductase domains of type I PKSS such	
FT	as eryA from <i>Saccharopolyspora erythraea</i>	
FT		927..19874
FT	CDS	

FT	/product="Sora
FT	/note="gene product is highly homologous to
FT	type I PKSS that are known to be involved
FT	in the synthesis of polyketide compounds"
FT	0.0. Title

FT	/note= "module 1 of SORA"
FT	7203.12884
FT	misc_feature

FT	/note= "module 2 of Sora"
FT	13455.19616
FT	misc_feature

```

FT      /note= "module 3 of SORA"
FT      19871..46318
FT      misc_feature

```

FT	19871.46318
FT	/*tag= g
CDS	

FT /note= "gene product is highly homologous to
ET type I PKS genes"

```

/*tag= h
/notes= "module 1 of SorB"

```

```
misc_feature 24638.30820
ET           /*tag= i
ET
```

[illegible]

PF	14-DEC-1996;	9605-0764233.
XX		
PR	24-AUG-1995;	93MO-US07954.
PR	08-JUN-1994;	9405-0258261.
PR	09-OCT-1996;	9605-0729214.
XX		
PA	(NOVS) NOVARTIS FINANCE CORP	

21 Beck JJ, Hill DS, Ligon JM, Neff S, Ryals JA, Schupp T
XX
DR WPI; 1998-158369/14.

PT DNA encoding Sorangium cellulosum polypeptide(s) - used for, e.g.,
PT biosynthesis of soraphen useful as antimicrobial agent against
PT phytopathogenic fungi

PS Claim 2; Columns 47-90; 64pp; English.

CC The present sequence contains a cluster of genes that encode polyketide
CC synthases (PKSs) that are involved in the synthesis of scoparins in
CC *Sorariaum cellulosum*. The proteins encoded by the present sequence are
CC *SoR1*, *SoR2*, *SoR3* and *SoR4* and *SoR3* contain biosynthetic modules
CC which contain a beta-ketoacyl synthase, an acyltransferase, a
CC ketoreductase and an acyl carrier protein domain, as well as beta-ketone
CC processing domains. *S. cellulosum* scoparins are useful as a cytostatic
CC and antimicrobial agent active against phytopathogenic fungi.
CC Scoparone-producing transgenic plants or biological control agents can
CC also be produced, which may reduce crop losses and nutritional
CC deprivation for local populations in many parts of the world.
XX
SQ Sequence 49377 BP 7247 A; 19522 C; 14477 G; 8131 T; 0 other;

alignment_scores:	
quality:	493.00
length:	2095
ratio:	0.593
gaps:	1022
percent_similarity:	39.666
percent_identity:	22.291

```

alignment_block:
US-09-508-821B-6 x AAV05287 ..
Align seg 1/1 to: AAV05287 from: 1 to: 49377

```

```

      62  GVALAGLYTHHROSSEGIYTHRIAAIAAIAAIAAIAA.....AI 75
      |||||.....|||
23912  GCGCGCGACACCAAGCGGCTACGCGCGCCCAACGCTCTCCGACGC 23966
      |||||.....|||
      75  AASPLYTRYHISATGGLYSERYLSALAEUFRTHGLHGLH..... 89
23962  CCGTGGCGACACCTGGCCACAGCGTGATGACACGACGCTGGTGAGCT 24011

```

90 ..GlyLeuGlnGlyArgProAlaPheProGlyTyrGlyValGlnAspSe 105
||| ||| : : : : :
24012 GGGCGCGTGCGGGCGGGCGCATGCCACCGCATAGCGGCAGCCACC 24061
||| ||| : : : : :
105 rSer.....ProTyrProGlyArgTyrAlaGlyGluG 116
||| ||| : : : : :
24062 CTCACAACAGCGGGCTGTGCGGATGCGCCCTTCGCTGCCCTGGCGGC 24111
||| ||| : : : : :
116 luserLeuGlnAlaIrpGlyAlaProGlnProProProGlnPro... 131
||| ||| : : : : :
24112 GCTGCGCTGCTCTGGAGCACGACGACGACGACGACGCTCGCGGACA 24161
||| ||| : : : : :
132GlnProLeu...ProAlaGl 137
||| ||| : : : : :
24162 TCGACTGGCGCGCTTTGCGCTTCGTCAGCGCGCTCGCCCGGCC.. 24209
||| ||| : : : : :
137 yValAlaLysTyrAspGluasnLeuMetLysThrAlaValProPos 154
||||| : : : : :
24210CGTCTCGCGGATTGCGCGA 24231
||| ||| : : : : :
154 eArGlnTyrAlaGlnGlnGlnValProPheArgThrHisSer 170
||||| : : : : :
24232 GCGCAGCGCGCTCTCGAGACCAAGAGGCG.. 24263
||| ||| : : : : :
171 LeuHisValGlnProProProGlnProGlnProLeuAlaTyrProLy 187
||| ||| : : : : :
24264CGTCTCGAGCATGGCGCGCC.. 24287
||| ||| : : : : :
187 sLeuGlnArgGlnLysLeuGlnAsnApLeaSerProLeuProPheP 204
||| ||| : : : : :
24287 24287
||| ||| : : : : :
204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSer 220
||| ||| : : : : :
24288CGCACTCTCTG 24299
||| ||| : : : : :
221 ThrTyrSerSerValGlnGlyGlyGlnGlyGlnAlaHisSerTyrLy 237
||| ||| : : : : :
24300 ACAAGCTCGGAGCGCTCGGAGAGCGAGCTGCTGCTGCTGCTG 24349
||| ||| : : : : :
237 sSerCys.....ThrAlaProThrAlaGlnProH 247
||| ||| : : : : :
24350 CTGGTGGCCACGAGAGCGCCCTGCTCTCGGCAAGAGCGCTCCCA 24399
||| ||| : : : : :
247 isaspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnAtgVal 263
||| ||| : : : : :
24400 TGTCGACCCCGAACAGGGCTTCTCGATC.. 24428
||| ||| : : : : :
264 GlnasnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280
||| ||| : : : : :
24428 24428
||| ||| : : : : :
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnSerArgH 297
||| ||| : : : : :
24428 24428
||| ||| : : : : :
297 ishIsalaGlnGlnThrLeuHisTyrGlnasnLeuAlaLysTyrGlnHis 313
||| ||| : : : : :
24429TCGCTCTCGATTGCTCA 24446
||| ||| : : : : :
314 TyrGlyGlnGlnGlyGlnGlyTyrCysGln.....ProAspAlaLaVa 328
||| ||| : : : : :
24447 TGGCGGTGAGCTTCGGCGCTTGACACAGCCACCGCATCAAGCTC 24496
||| ||| : : : : :
328 lArgThrProGluGln.....TyrTyrGlnThrPheSerP 340
||| ||| : : : : :
24497 CCGGCCACCTCGCTCGACCATCCCTCCTCATCGAGTGGCGCTCT 24546
||| ||| : : : : :
340 roSerSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
||| ||| : : : : :
24547 CTTGCGGAGCTGCTCGCCACCGCTCGGACGAGCGCTCTCGGCG.. 24593
||| ||| : : : : :
357 SerSerThrProSerProLeuMetProasnLeuGlnAsnPhe....ProTy 372
||| ||| : : : : :

```

25404 TCTCGCCCTCTGTCGGCGGCACGCCCATCAACAAGAGCGCGTCCGAGC 25453
625 a.....TPPAlaGluaIAPROS 631
| .....:|||||||+
25454 GGTATCACCGCCCCCAAGGAGCACTCCACCAGAGAAGTGCTCCGCGCGCC 25503
631 erLeuValLysAspSerSerLySProPhoPhSerLeu.....643
::: ||||| ||||| |||
25504 GCTCACACGACCCCAGCATACACCCCGCCAGCATCGAAGTGTGAGTCC 25553
644 .....GLuASnHisSerAlaCysLeuAspSerVa 653
25554 ATGACACCGGACACTCTCTGGGAAGAACCCATGAGATGC..... 25592
653 lAlaLysSerAlaTrpProAlaTrpProGlyLeuProAlaLeuProAsps 670
::: ||||| ||||| ||| |||
25593 .....AAgCCCHGGCG...CGATGTAGCGACGCGAGAACCG... 25628
670 erLeuGlInLeuAspLysGlyGlysnAlaLysAspRhpSerProGlyLeu 686
25629 .....CTGAAAAGCGCTCTCTCTGTGGCG 25652
687 pheGluAspProSerVaLaIAPheAlaThrProAspProLysThr 703
::: ||||| :||| ||| ||| ::
25653 CGCTCAGACACAAATAGTGGCATCTCCAGAGCGCGCTCCGGCGTCCGGCGGCC 25702
703 rGlyProLeuSerPheGlyThrLysProThrIleuGlyValProAlaProA 720
:::||| || |::: |||||::|||
25703 GTCCGCAAAATGG.....TGGCTCTCCGTCCGCAATGACGCGCCGCCGCC 25746
720 sPrOthrThrAlaIAPheAspCysPhePro.....AspThr 732
:::||| ||| ||| ||| |||||
25747 CACCTTCACACAGGCGCGCGGCAT....CCCTGATTGATTGGGATAACA 25792
733 ThrAlaIAserSerAlaAspSerAlaAspProPheAlaTrpProGlu 749
25793 CTCCGATGAGCGCTGTGATTAAC.....CCGAGGCTGTGGCGCGCCA 25836
749 uAsnLeuGlyAspAlaCysProAlaTrpGlyLeuHisProGlyLeuTr 766
::: ||||| :||| |||
25837 CGAA.....GATACAGTCCCGCC..... 25855
766 hLysGlyLeuGlnGlnGlyLysAlaSerAspArgLysLysGly 782
:::||| ||||| :||| |||
25856 .....CGCGCGGCGTCCGCGTTAGGAGATCTCCGCGAC 25891
783 AspThrHis.....GluAserAlaCysLeuGlyPheGlnI 795
:::||| :||| ||||| |||
25892 AACGCCACATCATCTCTGAGAGAGGTCCCGCGCCCTG.....TC 25932
795 uGluAspProGlyGlyLysValAlaSerLeuProGlyAspPheLys 812
::: ||||| :||| ||| |||
25933 GGCGAGCGCCGCAACCTCACAGAGAGCGTGGCGACGCTCCCGCGCGCT 25982
812 InGlnGlyValGlyValLysGlnGlnAlaGlyLeuLeuGlnCys 828
::: ||||| :||| ||||| |||
25983 GTGCGCTGCTCTGTCGGCGCAGAGAGCGCGCGCTCCGCGCCAGGCG 26032
829 ProGlu.....ValAlaLysAlaAspArgTrpLeuGlnAs 840
:::||| :||| ||| ||| |||
26033 AAGCGGCTCCGCGACACACTCTTGCCACAGACACTCGCCCTTANCGA 26082
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854
| .....:||||| ||| ||| ::
26083 TTGGCGCATATCCGACAGCGACACACCGCGCCACTTGAGACACCGCGCG 26132
855 LeuLeuProThrThrSerArgLysGlyAspLeuGlnAlaGlnGly 871
||| ||||| :||| |||
26133 CTCTCCGTGCGCGCGCGACCGCAGC..... 26155
871 rSerSerLeuCysGlnLeuLeuGlySerProGlnAlaTrpProGlyMetG 888
||| ||||| :||| ||||| |||
26157 .ACTCTCTCTCCGCGGTGATCTGCTGCTGCCACAGGACAAAGCCG..... 26199

```

26191CCCCGACCGCTTCGCGCGGAGCGGAGACCGACG 26234
26198 LnsPProLnsSerProLysAlaProLeuLleCysThrLysGlnGluVal 904
26235 GCAGAGCTGCTTCGCTTCCTCCGCGAAGGCTCCGACATGGGAAGGAGAT 26284
26285 GCGCTTCCTCCGCTCCAGCTCCGCGCGGATCCGCTCCGCGCTCAGCTCAGC 26334
26334 hrValGlnThrGlnSerLysValGlnSerThrPheGlnSerLysSer 947
26335 ATGGAGAGGCGCGCTCCGCTCCACAGTGC..... 26363
26364 HisMetLysProGlnGlnGlnLysProAsp.....GlyGlnAlaAlaPr 962
26364AGTGAAGCTGCTCCGCGCTCCGCGCGCGACGACGAGGAGGCC 26404
26364 OGlyAspSerThrThrSerAspAla..... 970
26405 CCCTCCCTCCGCGCGCGCTCCAGCTGCTACAGCCGCCCTTTGCCGTAT 26454
26455 GGTCTCCCTGCGCGCGCGCTCCGCGCGCTCCGCGCGGTCCGAGCCGCCGCCG 26504
26455 ProGlnAlaProIleAlaLysLysGlnLysProValProArgGlyLysSerLe 998
26505 TCGTGGCGCAGACGACGAGGAGAGATCGCGCGCGCTTCGCTGCGAGAGCCT 26554
26505 uArgSerArgArgAlaHisArgGlyLysProGluAlaGluAspSerProC 1015
26555 CTCTCCCTCCGAGACGCGCGCGCGCATGCCG.....CCCT 26589
1015 ys.....ArgAlaProValLysProLysAspLeuLeuProGlu 1028
26590 GCCCAGAAAGCGCTCACACGCTCCGCGCGCAGCGCGGAGTGGCGCGCG 26639
1029 SerCysThrGlyProProGlnGlnLysGln.....MetGlnLysAla.. 1041
26640 TCAGAGCTCGGCGCTTCGACCTCCACACTTACCTTCCTCCCTGGGGCGAC 26689
1042GlyAlaProGlyArgGlyAlaSerGlnG 1051
26680 AGGCTCTCCACCGCGCGCGCGACACAGCCACCGAGCTCACTCCGTATCCG 26739
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProArg 1067
26740 CGAGCCCGCGCGCGCGCTCGACGCGCTCGACGCTCCACCGCAACCAAG 26789
1068ThrProGlyProProGlyLysLeuThrThrThrProAlaPro..... 1080
26780 TGTTCGCGCGCAAGTCCGCTGCAGTCACTGACCTCCCACTCCGCGCAGATG 26839
1081ProAspLysLeuGlnG 1086
26840 GAGCGCGCTCAGAGAGAGACTCGCGCGAGCTCAGACCAACATCGCTCTCG 26889
1086 LysLysGlnAlaGlnAlaLysSerGlyLysArgValGlyLysProSer 1102
26922 CCGGACACAGGCTCGACGAGCTCCGAGCTCGACGCGCGCAGTGTATCGA 26971
1114AlaLeuProValAlaSerAspSerProMetGlnLysSerLysT 1128
26972 AACCTCGGGAACGCTCTGTTTCGAGGCGACACCGAGCGGCTCTTCGA 27021

[illegible]

```
257 CAGCAGCGACTGCCGCTGCCAGGCTGTGGCAGGGGGCCCACTCTCTAY 306
244 ..... 244
307 AAGAGYTCACAGACCCRNCTGCCAGCCCTCATGATAGCGCGATGATGC 356
245 ..... GlnProHisAspArgp 250
357 CAATGGGAACCTGGCTCCAGGCAACGGGTCCAGCAGCCCATCATCAGAGC 406
250 roleuThraLaserSerSerLeuAlaProGlyGlnArgValGln. .... 264
407 CGCTGACTGCCAGCTCCAGCTGCGCCCGGGGCGAGCGGTCCAGGGGCA 456
264 ..... 264
457 GCGGGTCCAGAGCCYCATGAYAGCCGCTGAGTGCARFKCSARCTGG 506
264 ..... 264
507 CYCRRGGGCAACGGGTCCAGATCTTACAGCTTACAGCCTGGCGGCTT 556
265 ..... AsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGln 278
557 GGCTAGCAATCTTCATGCTACCATGCTGGCGGCGCTGAGCTATGACAG 606
279 GlnGlnGlnGlnGln. Gln. .... 284
607 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 656
284 ..... 284
657 GCCTYRGCTAYGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 706
284 ..... 284
707 ACTTACAGCGCTCCACACACCCAGGAACACTCCACAGCAGCAGCAGC 756
285 ..... Gln 285
757 AGCAGAGCCCTTCAGAGCGGCGACATGCCAGGAACCTCCATCAG 806
286 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 302
807 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 856
302 rLeuHisTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlnGlnGlnGly 319
857 MCTCCATTACCAAGACCTCGCAAGTACCAACACTATGGACAGCAAGGCC 906
319 GlnGlyTyrCysGlnProAspAlaAlaValArgThrProGlnGlnTyr 335
907 AGGGTACTCTCCACTACCAAAACC. .... 932
336 GlnThrPheSerProSerSerHisSerProAlaArgSerValGlyAr 352
932 ..... 932
352 gSerProSerTyrSerSerThrProSerProLeuMetProAsnLeuGlu 369
933 .TCGCCAAGTATCAGCACTACGGCAGCAAGCAAGCCAGGCTACTGCCAGC 981
369 snPheProTyrSerGlnGln 375
982 AGCAC...TACGGGCAAGCA 998
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NAL1986.DAT.AAT06769

seq_documentation_block:

ID AAT06769 standard; DNA; 28598 bp.

XX
AC
XX

AAT06769;

```
DT 15-OCT-1996 (first entry)
XX Sorangium cellulosum soraphen gene cluster.
DE XX
XX Antipathogenic substance; soraphen; phenylpyrrole; antibiotic;
KW fungicide; pesticide; myxobacterium; ss.
XX Sorangium cellulosum.
OS WO9533818-A2.
XX PN 14-DEC-1995.
XX PD 30-MAY-1995; 95WO-IB00414.
XX PF 08-JUN-1994; 94US-0258261.
XX PR (CIBA ) CIBA GEIGY AG.
XX PA Beck JJ, Gaffney TD, Hammer PE, Hill DS, Lam ST;
XX PI Ligon J, Ryals JA, Schupp T, Uknes SU;
XX PT WPI; 1996-040226/04.
XX PS Claim 3; Page 140-158; 190pp; English.
XX CC This is the soraphen gene cluster from S. cellulosum. The gene
XX cluster may be expressed recombinantly to produce soraphen, or
XX expressed in a transgenic plant for disease-resistance.
XX SQ Sequence 28598 BP; 4031 A; 11703 C; 8263 G; 4601 T; 0 other;

alignment_scores:
Quality: 487.00 Length: 2135
Ratio: 0.553 Gaps: 121
Percent Similarity: 41.265 Percent Identity: 22.670

alignment_block:
US-09-508-821B-6 x AAT06769 ..
Align_seg 1/1 to: AAT06769 from: 1 to: 28598

18 GlnGlnThrSerGlnGlnThrSerArgLeuGlu.....As 29
22787 CAGGGAACCTCCAGACCTGCTGCTGCTCGACCCCTCGACGCTTGTGTC 22836
229 nTyrArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuL 46
22837 CTTTCTGCTCATCGACGCTCTGCGGCGCGGCAACAG..... 22879
46 euAlaLysAspTyrTyrAsnProGlnProTyrProSerTyrGlnGly 62
22880 .....CCGATACGCGCGG.....GAAACGCT 22903
63 AlaGlyThrProSerGlyThrAlaAlaValAlaAlaAspLysTyrHi 79
22904 TCTCGAGCGCTCGCGACCGACGCGGCTTGGACAGCGCGGACAG 22953
179 s.....ArgGlySerLysAlaLeuProThrGlnGlnGlyL 91
22954 TCGTGTGTGGGCGCGGTGGGCGGTGGTATATTATTCGCGGCG. 23002
91 euGlnGlyArgProAlaPheProGlyTyrGlyValGlnAspSerSerPro 107
23003 .....CCCTGGCAGCCGACGCTGGACACACGCTGCTGTGCGCGA 23041
108 TyrProGlyArgTyrAlaGlyGluSerLeuGlnAlaTyrPglyAlaPr 124
23042 TGGCCCTTCGCTGCGGCGGCTCGCGCAAGCCCTGGAGCAGCAG 23091
```

```

124 oGlnProProProGlnPro.GlnProLeuProAlaGlyValAlaLys 140
    :::::::::: :::::::::: :: :: :: ::::::::::
23992 GAGACCAACCGTACCGCTCCGACATCATGAGCGGCCG...CTTGGGCC 23138
141 TyrAspGluAsnLeuMetLysLysThrAlaVal.ProProSerArgGlnTyr 157
    :::::::::: :::::::::: :: :: :: ::::::::::
23139 TTCGATTCAG.....CGTCCGCTCCCGCCGCTCTCCGCGACT 23179
157 rAlaGlnGln.....GlyAlaGlnVal.ProPheA 167
    :::::::::: :::::::::: :: :: :: ::::::::::
23180 GCCCGACGACGCGCCCTCGAAGACAGAGAGCGCG... 23216
167 rGThrHisSerLeuHisValGlnGlnProProProGlnGln..... 181
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23217 ..TCCTCCCGACACAGCGCGCGCGCGCGCGCTCTCGACAACTCCG 23264
182 ...ProLeu.....AlaTyrProLysLeuGlnArgGlnLysLeuG 194
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23265 GAGCCGCTCGAGAGCGAGCGAGCTCGCTCTCGCGCGCGCTGCTCG 23314
194 nAsnAspIleAlaSerProLeuProPheProGlnGlnTyrHisPheProG 211
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23315 ACAGAC.....GCCCTCGCTCTCGCGCGCGA..... 23342
211 InHisSerGlnSerPheProThrSerSerThrTyrSerSerSerValGln 227
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23343 .....AGCGCGCTTCCGACGCTCGAC..... 23363
228 GlyLysGlyGlnGlyAlaHisSerTyrLysSerCysThrAlaProThr.. 243
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23364 ..... 23371
244 ...AlaGlnPro.....HisAspArgProLeuThrAlaS 254
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23372 GCTTCTTGACCTCGCTCGATTCGATTCGACCT.....CGAGCTT 23415
254 SerSerLeuAlaProGlyGlnArgValGlnAsnLeuHisAlaTyrGln 270
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23416 CGTCCGCGCTTCGACAGCGCGCGCG..... 23441
271 SerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGlnGlnGlnGln 287
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23442 .....CATCAAGCTCCCGCGCGCGCC 23461
287 nGlnGlnGlnGlnAlaLeuGlnSerArgHisAlaGlnGlnGlnGlnGln 304
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23462 TCGCTTCGACATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 23502
304 lAsTyrGlnAsnLeuAlaLysTyrGlnHisTyrGlyGlnGlnGlnGly 320
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23503 ATGC..... 23506
321 TyrCysGlnProAspAlaAlaValArgThrProGlnGlnTyrTyrGlnTh 337
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23507 .....GGACTCGCTCGCC..... 23521
337 rPheSerProSerSerSerHisSerProAlaArgSerValGlyArgSerP 354
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23522 ...ACGCCCTGGCAGAGAGCTCTCGCGAGCGAGCGCGCGCGCTC 23567
354 roSerTyrSerSerThrProSerProLeuMetProAsnLeuGlnAsnHe 370
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23568 CGCGCGCGCTCGAGAGAGAGAGCGCA.....TCG 23596
371 ProTyrSerGlnGlnProLeuSerThrGlyAlaPheProAla..... 384
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23597 CCATCGTCGCAATGAGCTCGCGCTCGCGCGCGCGCTCGCGCGAGTTCGAC 23646
385 .....GlyIleThrAspHisSerHisPhe.. 392
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23647 GCTCTTGGAGATTCCTCAACCAAGGCGCGAGCGGCTCGAGCCATTC 23696

```

```

393 .....MetProLeuLeuAsnProSerProThrAsp 402
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23697 ACAGACGCGCTGGAGACCGCGCTCTCTAGACCCCGAGACCCGAGCG 23746
403 .....AlaThrSerSerValAspThrGlnAlaGlyAsnCysGlyAsP 416
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23747 ACGCCAAGAGCTACGCTCGGCGATCGCG.....CGATGCTCG... 23782
416 OleuGlnLysAspLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuG 433
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23782 ..... 23782
433 InsertLeuThrAlaLeuThrLeuGlnValGluAsnIleSerAsnThrVal 449
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23783 .....ACCAATCGAC 23794
450 GlnGlnLeuLeuSerTyrAlaAlaValProGlnLysGlyValLys 466
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23795 TCTTCGACCTGCTCTTCGCGATTCAGCGCGCGGAGAGCGCAAC..... 23839
466 sAsnLeuValSerArgThrProGlnGlnHisLysSerGlnHisCysSerP 483
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23840 .....ACCTCGACCGCC.....AGCACCGCGCTGCTCT 23867
483 roGlnGlySerGlyTyrSerAlaGlnProAlaGlyThrProLeuSerGln 499
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
23868 CGAATCTCCCTGCTG.....CCCTCGAGAGCGCGCGCGATCTCTC 23908
500 ProProSerSer.....ThrProGlnSerThrHisAla..... 510
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23909 CCACCTCCCTCAAGACTCCCTCAAGCGCGCTCTCTCGAGCTTCGCGC 23958
511 .....GlnProGlnGlnAlaAspTyrLeuSerG 520
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
23959 GCGCAATACGGAGTCAAGAGCGGAGCTCGGAGAGTTCGAGATTACTT 24008
520 LysSerGln.....AspProLeuGlnArg...SerPheLeuTyrCys 532
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
24009 CATCAAGAGCACTTCGCGCTCTTCGCGCGCGCGCGCTTCGCTTACGC 24058
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLys 549
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
24059 TCGGCGCTCGAGGCGCGCGATCTTCG.....TCGACACGCGCTCTCC 24102
549 sProGlnSerValSerThrCysSerValThrSerProAspAspMetSer 566
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
24103 TCCGCGCTGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 24152
566 hLysSerAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
24153 GTCGAAGCTCGCGCTCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCT 24202
583 SerPheSerLysPhe.....ValAlaGlyI 591
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
24203 CCTTCGTCATCTCTTCGCTCTGCGCGCTTCGCGCGCGAGCGCGCTCC 24252
591 uArgAspCysProArgLeuLeuLeuSerAlaLeuAlaGlnGlnAspLeuA 608
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
24253 AAGACCT..TCTCGGACACGCGGACGCTTCGAGACGCGGAGAGCGCT 24289
608 lAserGlnIleLeuGlyLeuGlnGlnAlaLe.....GlyGlnLys 621
    :::::::::: :: :: :: :: :: :: :: :: :: :: :: ::
24300 CCTTCGCTTCGCGCTCGAGCGAGTCCGCGCGCGCTTCGCGCGAGAC 24349
622 AlaAspLysAlaThrAlaGlnAlaProSerLeuValLysAspSer..... 636
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
24350 GCGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 24399
637 .....SerLysProProPheSerLeuGlnAsnHisSerLysLeuA 651
    :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
24400 AGCGGTATCACCGCGCGCGACGACCTCCACGAA..... 24436
651 sPheSerValAlaLysSerAlaThrProArgProGlyGlnProGlnAlaLeu 667

```


24437AGTCTCTCCGGCGCGCTCCACACCCCGCATCACCC 24475
668 ProAspSerLeuGlnLeuAspGlyGlyValLeuLeuGlyProThrValG1 933
24476 CGCGCGAGC.....TCGACGTCGTGAGTGCATGGCACCGGACCC 24516
684 oglyLeuPheLeuAspProSerValAlaPheAlaThrProAspProLysL 701
24517 T...CGTGGGAGACCCATCCATGAGTCCAAAGCCCTGGCGCGCTACAGC 24563
701 ysThrThrGlyProLeu.....SerPheGlyThrLysPro 712
24564 CGACGCGACACCGCTGAAGACCTCTCTCTCGCGCGCTCAAGACCA 24613
713 ThrLeuGlyValProAlaPro..... 719
24614 ACATCGCCCATCTCGAGCGCGCTCCGCGCTCCGCGGCTCGCCAAATG 24663
720 AspProThrThrAlaAlaPheAsCysPheProAspThrThr..... 733
24664 GTGGCTGCTCCGCGACACCGCTGCGCGCGCTCCGCGGCTCCAGCGACCC 24713
734AlaAlaSerSerAlaAspSerAla..A 742
24714 ACGCAATCCCTCATCGAGTGGGAGCGCTCGCCATCGACGTCGTGGATA 24763
742 snProPheAlaThrProGluGlnLeuGlyAspAlaCysProArgTrp 758
24764 CCGCGAGCGCTTGGCGCGCGCGACAA.....GATGCGAGTCCCGCG... 24804
759 GlyLeuHisProGlyGluLeuThrLysGlyLeuGluGlnGlyGlyLysAl 775
24805CGCGCGCGCATCTC 24818
775 aserAspGlyLeuSerLysGlyAspThrHis.....GluAlaAs 788
24819 CGCTTCGATCTCGGCGACCAACCGCGCTCATCTCGAAGAGGCTC 24868
788 erAlaCysLeuGlyPheGlnGluAspProGlyGluLysValAla 804
24869 CGCGCGCGCTG.....CGCGCGAGCGCGCACCC 24897
805 SerLeuProGlyAspPheLysGlnGluValGlyGlyVal..... 818
24898 TCACAGCGCGCGCGCGCGCGCTCCGCGCGCGCGCTGCTCTCTCTGTC 24947
819LysGluGluAlaGlyGlyLeuLeuGlnCysProGlu..... 830
24948 GCCAGGAGCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24997
831ValAlaLysAlaAspArgTrpLeuGluAspSerArgHisCysCys 845
24998 ACCTGTCG 25047
846 SerThr.....AlaAspPhe.GlyAspLeuProLeuLeuProThrS 860
25048 GCCACACCG 25097
860 erArgLysGluAspLeuGluAlaGluGlnLysSerSerLeuCysGlu 876
25098 CGCGGAG.....AGCTCTCTCCGCGC 25120
877 LeuLeuGlySerProGluGlnArgProGlyMetGlnAspProLeuSerPr 893
25121 TCGACTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25149
893 oLysAlaProLeuLeuLysThrLysGluGlnValGluGlnValLeuAsps 910
25150 CCGAGACCG.....TCTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 25190
910 erLysAla.....GlyTrpLysSer 916

1154 uilepheHisSerLysArgArgProSerGlu..... 1165
1166GlyArgLeu...ProAsnGlyAlaThrLysLysLeuAsp 1179
1180 AsnSer.....HisLeuProAlaThrPh 1187
1187
12619 CGGCGCCCTCGCTCCCGCAGAGTGTCCCTCCACCTACCTCCCTCCGAC 26218
1187 eLysValSerSerSerProGlnLysGluY...ArgValSerGlnArgA 1203
12619 GCGAGCGCTTCGTGCTGAGCGCTCCAGCGGCAGCTGCGCAGACGTCC 26268
1203 LaArgValProLysProGly.....AlaGlySerLysLeuSerAspArg 1217
12629 TCCTCCAGGCTTACCTGCGCCGACACCGCGCTGCTCGCGCGCTGCG 26318
1218 Pro.....LeuHis.....AlaLeuLysArgLys 1225
126319 CCGCGCGCAGCGCGATGCGCTTGTCTTCACAGAGCGCTCTCCCTGCGAG 26368
1225 sSerAlaPheMetAlaProValProThrLysArgAsnLeuValLeuA 1242
126369 A.....GCACCCGTGGCTGGAAGACCAAGCTGT.....CTTC 26400
1242 rGHisGlySerSerSerSerSerSerSerSerSerSerSerSerSerSer 1258
126401 GGCAATACCTGTCTC..... 26414
1259 ArgArgArgGlyLeuArgValProProProSerSerArg...GlyCysLe 1274
126415TGCCAGCGCGCGCTCTCCAGAGCGCTCTGCTGAGTGC 26452
1274 uLeuProArgLysProSerProProArgAlaMetAlaSerLeuProGlns 1291
126453 CCCATCTGTGTGGCTCGACACCTCGAAGACGTCAAGACGTGACACCCGCC 26502
1291 erSerHisPro.....ArgArgProProPheLeuProGlnAla 1303
126503 CTGCGCTGCCCATCGCAGAGGCGCGCTCTCTCCATATCTCGTGGGCC 26552
1304 ArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysValLeuProP 1320
126553 CGCGGACGCTGCTGAGCAGAGGCGCTCTCTCTCATACCG...GCGCC 26599
1320 ArgGlySerLysArgLysLeuLysLeuGlnAlaLeValGlnLysIleThr 1337
126600 ACGACGCGCTTCAGAGAGGCCCTCTGAGCTGACACCGCCGCGCTCTGC 26649
1337 er.....ProSerLeu..... 1340
126650 GCGCAAGCTAGCCCGTCCATTCGCTTCGATGCTCCGCGATAGGCCCCC 26699
1341Lys 1341
126700 CTGCGCGCCACACCGAGGTGAGACCCAGAGTTTCTACGAGCCCTCGAGA 26749
1341 sLysPheAlaCys.....LysAlaProGlnLysAlaSerProGlnAsn 1354
126750 GCGCTGGCTGTGCTTATGCCCCCGAGTTCAGGGCTCTCGCCGCGCTGTA 26799
1354 1354
126800 CAAGCGCGCGACAGAGCTCTTCGCGAAGCAAGCTCCGAGCGCGCG 26849
1355ProLeuSerProSerLeuSer..... 1361
126850 AAGAGACGCGCTCGTTTGTCCCTCCACCGCGCTGCTGACAGCGCC 26899

1362AspLysAspArgGlyLeuL 1368
126900 TTGAGGCGCTGCGCTTTGTAGACGACACGAGGAAGGCGCTTCAGATGCG 26949
1368 ySGIAlaGlyLys.....SerProValGlyVal..... 1377
126950 CTCTCTGAGAGGAGATGCTGCGCTCGCGCTCGAGACACCACTCTG 26999
1378GluGluGlyLeuValAsn.ValGlyThrGlyG 1388
127000 CGCGGCGTTTCCACCGCTCCAGAGGCGAGATCCCTCGCGCTCTCTCT 27049
1388 LysLysLeuProThrSerGlyAlaAspProLeuGlyArg..... 1400
127050 CGCCGACGCGAGGCGAAGCCATGCGCTCTGCTGACAGCGTCCGAC 27099
1401AsnProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLys 1415
127100 GCGCGCGCTTCGCGAGAGCTCGCAGACCGCGAGCGCGGAGCTCCAC 27149
1415 sLysLeuSerSerThrAspCysPheLysThrGlnAlaPheThrSerProG 1432
127150 TGCCCTCTTCGCGATGCGACTGAGGAGCGAGCGCAAGCCCGACCTGAC 27198
1432 LysLysLeuGlnProGlyLysThrAlaLeuAlaProLysLysArgSerArg 1448
127199CCCATGCGCGAGGCTGCTCTCTC..... 27225
1449 LysGlyArgAlaGlyAlaHisGlyLysLeuSerLysGly.....ProLe 1462
127226GCACAGAGAGTCTGACCTCGGAGACCGAGGAGTCTCT 27263
1462 uGluLysArg.ProThrLysGlyProAlaLeuLeuLeuThrProArgAsp 1478
127264 CGACCGCTATACCGACCTTG.....CTGCTTACGAGAGCGCTCGAC 27307
1479 ArgAlaSerGlyThrGlnGlyAlaSerGlnAspSerSerGlyGlyGly 1495
127308 AGGCGCGCTGCGCTTCACAGGCTGTGATCGCCCTTCA.....TC 27348
1495 LysLysProLysMetGlnLysLeuGlyProAlaSerGlnProGluG 1512
127349 GCTCTGCGGAGAGGAGACCTCAGCGAGGCCCGCGAGACACCGCGCA 27398
1512 LysArg.....ProCysGlnProGlnThrArgAlaGlnLysGlnProGly 1526
127399 CGCGCTGCGCTTGCAGAGCTGTGCTGCGAGAGCGCGCTGCTCTCT 27448
1527 HisThrAsnTyrSerSer.....Tyr 1533
127449 CGCGCTGCGCTCTGCTTCACCGAGCGCGCTGCGCACCGCTGAGAA 27498
1533 rSerLysArgLysArgLeuThrArg.....GlyArg 1543
127499 GAGCTCAGAGGCGCTGCTCAGCGCGCTCTCTGAGGAGTCTCGCTGCGC 27548
1544 AlaLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArg 1560
127549 GCAGAGGAGACACCGAGAGCGCC.....TCTGCTCTGCTGAGACTTG 27592
1560 gGlnGlnGlnValLeuProLeuAspProAlaLysProGlnLysArgLeu 1577
127593 ACGACAGCGAGGCGCTCCAGACACCGCTGCGGCGCGCTGAGCG... 27638
1577 ySTyrlleSerSerCysLysArgLeuArgSer..... 1587
127639AAGAGACCGAGAGATCGCGCTCGCAGAGGCA 27671
1588AspSerArgThr 1591
127672 ACCCGCTGAGCAGCGCTCTCATCAGAGAGGACACCGCGCTGCGCG 27721
1591 rProAlaPheSerProPheValArgValGluLysArgAspAlaPheThr 1608

```
27722 CCTGGTCGGCGCCCTCTGCTGAACACGAGCG..... 27758
1608 hrileCysThrValAlaSerProGlyAspAlaProHisArg 1624
27759 .....CAAGACCTGCTCTCACTC 27779
1625 LysPro.SerSerSerAlaSerSerSerSerSerPheSerL 1641
27780 GCGCCAGGGCGGAGCGCTCGGGGCTGATGCTTGGAGC...GAGC 27826
1641 euAspAlaIaGlyAlaSerLeuAlaThrLeuProGlyGlySerIleLeu 1657
27827 TCGAGCTCTGGGGCTTGGTC...ACCTCGCGCGTGGAGAGTGGCC 27873
1658 GlnProArg.....ProSerLeu...Pr 1664
27874 GATCCACGCGCTCTAAGGACCTCTGTGATAACATTCCGAGCGTCAACC 27923
1664 oLeuSerSerThrMethHisLeuGlyProValValSerLysAlaLeuSerT 1681
27924 GATCGCGCGCTGCTGATCGCGAGCGCTCTCGAG..... 27960
1681 hrSerLeuValCysCysLeuCysGlnAsnProAlaAsnPhelLysAsp 1697
27960 ..... 27960
1698 LeuGlyAspLeuGlyProTyrTyrProGluHisCysLeuProLysLy 1714
27961 ...GGCGATCTCTCGCGCGCAGCTCTCGAGCGGAT..... 27995
1714 sLysProLysLeuGlyGluLysValArgProGluGlyThrCysGluGluA 1731
27996 .CGACCGCGCTT.....CGCCCCAAGATCGATCGCGCGCTGG 28032
1731 la..... 1731
28033 CACTGATCAGCTCACCAGATAGACCCCTTGGCGCTTCATCTCTT 28082
1732 .....SerLeuProLeuGluAr 1737
28083 CTCGTCGTCGGCGGCTCTCGGACGCTCAGGTCACTTCCAACTAGCGG 28132
1737 gThrLeuLysGlyProGluCysAlaAlaAlaAlaAlaThrAla 1750
28133 CTGGAGCGCTTCTCTCGATCGCTTGGCGACACCGCGG 28172
seq name: /SIDS2/gcgdata/geneseq/NA2000.DAT:AAA92302
seq_documentation_block:
ID AAA92302 standard; DNA; 31422 BP.
XX AC AAA92302;
XX 10-JAN-2001 (first entry)
DT S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
DE Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX OS Streptomyces avermitilis.
XX Key Location/Qualifiers
FH 1.14646
FT CDS /tag= a
FT /note= "avermectin aglycon synthase protein"
FT 14824..31422
FT FT /tag= b
FT /note= "avermectin aglycon synthase protein"
XX WO200050605-A1.
```

```
XX 31-AUG-2000.
XX 23-FEB-2000; 2000WO-JP01041.
XX 24-FEB-1999; 99JP-0046961.
XX (KITA) KITASATO INST.
XX Omura S, Ikeda H;
XX WPI: 2000-565458/52.
XX P-PSDB; AAB23751, AAB23752.
XX Avermectin aglycone synthase DNA and proteins encoded by all or part of
XX it for the production of avermectin and its derivatives for drug and
XX agrochemical use
XX Claim 2; Page 134-203; 314pp; Japanese.
XX The present sequence represents DNA which encodes avermectin aglycon
XX synthase proteins. Also described are: (1) polypeptides encoded by all
XX or part of the DNA; (2) expression vectors containing the DNA; (3) host
XX cells transformed by the vectors; (4) preparation of the polypeptides
XX by culture of the transformants; (5) preparation of avermectin aglycon
XX or its derivatives by culture of transformed avermectin-producing
XX microorganisms; and (6) oligonucleotides of 5-60 bases in length
XX containing sense or antisense sequences from the avermectin aglycon
XX synthase DNA. The enzymes are useful for the production of modified
XX forms of avermectin and of the intermediates in its biosynthesis, for
XX use as drugs, veterinary drugs and agrochemicals.
XX Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;
SQ
alignment_scores:
Quality: 470.50 Length: 2105
Ratio: 0.560 Gaps: 105
Percent Similarity: 39.905 Percent Identity: 21.235
alignment_block:
US-09-508-821B-6 x AAA92302/rev ..
Align seg 1/1 to reverse of: AAA92302 from: 1 to: 31422
22 GlnGluThrSerArgLeuGluAsnTyrArgGlnProSerGlnAlaGlyLe 38
6554 CGCAGGACGTCGAGGAGAGACAGTCCAGGTACGAGGCTTGGGCG 6505
38 uSerCysAsp.....ArgGlnArgL 45
6504 GCAGTCGACATCTCTGGCGGAGACCGGTGAGGAGCGAGGAGTTCT 6455
45 euLeuAlaLysAspTyrTyrAsnProGlnPro..... 55
6454 GCGCCTATGCTTCCAGTGGGTGCTCTCTCCGACAGCATGGCGATC 6405
56 TyrProSerTyrGluGlyAla.....G1 64
6404 TTTCGTCGCGCGCGGTGCGGCGCACACCTGTATCCACCCCGCGGTGG 6355
84 yThrProSerGlyThrAlaAlaAlaValAlaAlaAlaAlaAlaAlaAla 81
6354 TGCCTCTCGCGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6305
81 ly.....SerLys 83
6304 CGCCGAGGACCGACGATGCTCGACACCGCGCTCCGACACCAA 6255
84 AlaLeuProThrGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTy 100
6254 GCGCGCGGAGCGGAGTCACTCCACCCACACATCCGACCC..... 6210
```

```
100 rglvalglaspserprotyr..prolylarginylglutln 116
    |||
6209 .....CCACACACCGCCACACACCGCCACACCGCCG 6176
    |||
117 Serleuglnalatr.....glyalaprolinpro.....Pr 127
    |||
6175 CCGCCCCCGGACAGCCAACTCCGACCGCCGACACACCGACGAC 6126
    |||
127 oproprolinproglinproleuprolalaglyalalalyspeltua 144
    |||
6125 ACCCCGACACCGGACACACACCGCCGACACCGCCGACCGCCG 6076
    |||
144 spleumetlylsthralalaproproserargylmthralagln 160
    |||
6075 ACCCCCATCCGCGCCCGACGACGCTCTCCAAATCCATCCGATCC 6026
    |||
161 glyalaglnvalprophargthrhisserleuhsialaglnpropr 177
    |||
6025 TCCCACTCACCCGACAGACACCGCCGACCGCCGACCGACCTCC 5976
    |||
177 oproprolinproleuallatyrprolysleuglnarglnlysleug 194
    |||
5975 ACCCCGCGCCACTCCA..... 5960
    |||
194 lnamaspillealaserproleuprophneproglinlythrhisphero 210
    |||
5959 .....CCTCATCCGCA...ACACGACACG 5936
    |||
211 glomhisserglaserphneprothrserthrtyrsersevalgl 227
    |||
5935 ACCCACTCCACCATCCACCT..... 5915
    |||
227 ngllygllyglnglnlyalanhisertylsercysthralprothra 244
    |||
5914 .....GCACGACG 5905
    |||
244 laglnprohis..asparproleuthralasersevalalaprogl 260
    |||
5904 CTCATCCACATCCAGACGTCG..... 5884
    |||
260 yglargvalglasleuhsialatyrglinserylarginleusertyra 277
    |||
5884 ..... 5884
    |||
277 spglnglnglnglnglnglnglnglnglnglnglnglnalaleu 293
    |||
5883 .....CGGACACACCGCCGCG 5866
    |||
294 glinsertarghshsialaglnluthleuhsialtyrglnasleuallatyr 310
    |||
5865 CAACCCCATCCATCTGTATACACCGCCGACACCGCCGACGACCTCCG 5816
    |||
310 styrglnhisertylglnglnglnglnglnglnglnglnglnglng 327
    |||
5815 C.....ATGCCCATGTCTC 5802
    |||
327 lavalargthrproglnglnlytyr.....glthrpheserpro 340
    |||
5801 GACTTCACCGACCCACACACACCGCGCTCCCGGACGACCTGCTCC 5752
    |||
341 serseerhisserproalarg..servalgly..argserproserthr 356
    |||
5751 ATAAAGTGGAGACACGCTGCGCTCATGATGATACCAACGATCGTAC 5702
    |||
357 serseerthrproserproleuwerproasn...leuglnasphproty 372
    |||
5701 CCGTCCCGTGGCCCTCCACACATCCACATCCGCGACGACCGCCCA 5652
    |||
372 rserglnlnproleuserthrlyalapherproalaglyllethrasph 389
    |||
5651 CACGCCACGCGCTGCGCATCA.....CCGCGTACGCGACGACGAC 5611
    |||
389 ls.....SerHisphenet 393
    |||
5610 ATTGCGCGCGCTCAACCCATTCGACGACGCTCTCTGATTCACCGACCTCC 5561
    |||
394 proleuleuasproserprothrasp.....AlatHis 405
    |||
5560 CCGGACACACCGCCGACAAACCGATGACACGACGCTGACCTGCGACAGC 5511
    |||
405 rservalaspthrlnalaglyasnscylslyproleuglnlnasplysl 422
    |||
5510 CGCTCCACACACGACACACCGCCGACACCTCGGCGGACCGCCATGCG 5461
    |||
422 euprolglnsleuuseraspserleuinsertleuthralaleu 438
    |||
5460 CCGCG..... 5456
    |||
439 thrleuglnvalgluaslnleserasthrvalglnlnglnleuense 455
    |||
5455 .....AC 5454
    |||
455 rlysalalalavalproglinlyslslygllyalysasleuvalserargt 472
    |||
5453 CCGTACGCTTGCACCGCGCGCTCCGCGACACACCGCCGCTGCGACGAA 5404
    |||
472 hrproglnglnhislyserglnhislyserproglnglyserglytyr 488
    |||
5403 CTCACACACGACCGCGCGCTG...ACATCACG..... 5372
    |||
489 seralaglnproalaglythrproleuserglproproserseerthrpr 505
    |||
5371 .....TGACACCGCGCGCGCGACGACGACGACGACGACGACGACG 5325
    |||
505 oglnsertthrhisalaglnpro.....glnlnalaspityrleus 519
    |||
5324 TGACACGACGACGACGACGACGACGACGACGACGACGACGACGAC 5275
    |||
519 erglysercluasproleuglnargserphelautyrcysanglnala 535
    |||
5274 CGTACCGCGCGGACCT.....CAGGC 5252
    |||
536 arglyserproalargvalanser..... 544
    |||
5253 CAGAGGTGAGGACGACGCTCGGATGAGGACGCTGAGACGCGCGCTC 5202
    |||
545.....Anserlysalalysproglusertvalserthrtyrservalthrs 560
    |||
5201 ATGACGCTAGCCCTGAGCTCTGCGGCGCGCGCGCGCGCTG 5152
    |||
560 erproasparpserseerthrlyserasparsephe..glnserleu 576
    |||
5351 GTCGTACGACACGACGCGCGGACGACGACGCTGAGACCGCGACG 5102
    |||
576 sglysertleuproleuasparseerpheserlysphevalalaglyglua 593
    |||
5101 TGGGCGGCGTCAATGCCCGCGCTCCACACGCTCCGACGACCTCCAGC 5052
    |||
593 spcyspro.....ArgleuenseerAlaleuallaglnaspleu 607
    |||
5051 ATCAACCGCTGCTGCGATCCATCCGACGCTTACCGGATATCC 5002
    |||
608 Alasergluileuglyleuglnglnalalaglyglulysalasp..L 624
    |||
5001 GAAACACCGCGATGACCTGCGCGGACGCTCCGACGACGACGACG 4952
    |||
624 ysalatrpalaagln..Alaproserleuvalylasparseerlysp 640
    |||
4951 GCGATATGACGACACCGCGCGCGCTGCGGATCATGAAAGCGCTCC 4902
    |||
640 opheerleuglnasnhiseralacyseleuaspevalalalysera 657
    |||
4901 A..... 4901
    |||
657 latrpargproglnglnproglualaleuoproasparseerleuglnleu 673
    |||
```

4900 ..CGTCCACACCGCGGTGACCGGAAATCTCCACCGCAT..... 4862
674 AspLysGlyGlyAsnAlaLysAspPheSerProGlyLeuPheGluAspPr 890
4862 4862
690 oSerValAlaPheAlaThrProAspProLysLysThrThrLeuProLeuS 707
4861 4861
707 erPheGlyThrLysProThrLeuGlyValProAlaProAspProThrThr 723
4840 4840
724 AlaAlaPheAspCysPheProAspThrThrAlaAlaSerSerAlaAspSe 740
4804 AAGCGCAGCCCATCC...CCACATCGCAATCGCGTCGCAACGTCGACA 4758
740 rAlaAsnProPheAlaThrProGluGluAsnLeuGlyAspAlaCysProA 757
4757 CGCGAGCCCG...GGACCGGAGGAGCAATGTCAGCTGGCCGCCCGC 4714
757 rGtrpGlyLeuHisProGlyGluLeuThrLysGlyLeuGlu.GlnGlyG 773
4713 GC.....CCTCCTCCAGCTGTCTTGTAGGTATCCGGCCAGCGCG 4673
773 yLysAlaSerAspGlyLeuSerLysGlyAspThrHisLysAlaSerAlaC 790
4672 A.....GGAGTGGGTGTGCAAGATCGGT..... 4645
790 ySLeuGlyPheGlnGluAspProProGlyGluLysValAlaSerLeu 806
4644GGTGGCAGGAGG.....CCGGTGGCGGTGAGGGGT 4606
807 Pro.....GlyAspPheLysGlnGluValGlyCl 817
4605 GCGCAGTTCGACGGCGCTCAGCGAGCAAGCCAGGTCCGGAAGGCTC 4556
817 yValysGluGluAlaGly.GlyLeuLeuGlnCysProGlu...ValAla 832
4555 GCTCAGGGCTACGGGGTGGGGTGTG...TGTCGAGCAGCGTCGCC 4509
833 LysAlaAspArgTrpLeuGluAspSerArgHisCysCysSerThrAlaAs 849
4508 GCGTACGTACGACGAGTCGAGAGGCGACGTCCTCTCGCGGTGTC 4459
849 pPheGlyAspLeuProLeuLeuProProThrSerArgLysGluAspLeuG 866
4458 CATGCC.....TTGAGCCGTGCGGAGAACGAGT 4430
866 LuAlaGluGluGluTrpSerSerLeuCysGluLeuLeuGlySerProGlu 882
4429 CGGGGGAT..... 4422
883 GlnArgProGlyMetGlnAspProLeuSerProLysAlaProLeuLeuCy 899
4421 ...CGGTGGCGGTTCGAGTCCGGTGTTCGGGGAG...GGGTG 4379
899 sThrLysGluGluValGluGluValLeuAspSerLysAlaGlyTrpGlyS 916
4378 CTTCCGGGATGTCCTGATGAGGGCGGAGTGGGAGCCGGGAGGGA. 4330
916 erProCysHisLeuSerGlyGluSerValIleLeuLeuLysProThrVal 932
4329GTGGCGGTGATCGTCCAGTC.....GATGTC 4300
933 Gly.ThrGluSerLysValGlnSerTrpPheGlu..... 943
4299 GGGACGCTCACAGGTCCTGATGTCCTCAAGGCTGGCCAGTGCCA 4250
943 943
4249 CCAGCGCGCTCTCCGGGTCTATCGCGCCAGACCCGCGCATCTGC 4200

944 ...SerSerLeuSerHisMetLysProGlyGluGlnGlyProAspGlyG 959
4139 CCACAGCGCCCTCGCGCATCTCCCCACACCGCGGACCCACCGCCAC 4150
959 uArgAlaProGly.....AspSert 966
4149 GCGCACCCTCGCGCGCTCACCACCGCGGTGCGGAACGATTGCTCCA 4100
966 hrThrSerAspAlaSer.....LeuAlaGlnLysProAsnLys 978
4039 CATACGGGTTCGCGCGCTTAATCTCCCTGTCGCGCGCGCGACGTC 4050
979 ProAlaValProGluAlaProLleAlaLysLysGluProValProArgG 995
4049 GCGCAGCGCGAGAGACACACCAACCCCGAAGATCCGCCGCCCGCT 4000
995 yLysSerLeuArgSerArgValHis.....ArgLysLeuPro. 1008
3999 CACTCATGCAATTCAGGCGCGCGCGCTTCGCGCGAGCACCCCGG 3950
1009GluAlaGluAspSerProCysArgAlaProValLeuProLysAsp 1023
3949 TGACACGCTCGGAGCAACCTCTCCAAACCGCG.....TCA 3912
1024 LeuLeuLeuProGluSerCysThrGlyProProGlnGlyGlnMetGlu 1040
3914 TTCCAACTCCCGCGCATGACCAACACCGCGCGGCGACCTCCGCCG 3862
1040 y.....AlaGlyAlaProGlyArgGlyAlaSerG 1050
3861 ACAGCGCGCGCAACACTCCGCCAACGCTCAGGTCGCGCGCATCAC 3812
1050 LuGlyLeu..... 1052
3811 ACGCCACACCGACACCCGCGCGCCCAAGCCCATCAAGTCCGCTCGGAGT 3762
1053ProArgMetCysThrArgSerLeu.. 1060
3761 TCTTCGACTCCTCGGGGTCTCCCGCGGTGCGCTCACCAGCAGGTG 3712
1061ThrAlaLeuSerGluProArgThrP 1069
3711 TTCGGCGCACCGCGGCCATCCACCGGGGAGCTGCGCACCACTCGC 3662
1069 roGlyProProGlyLeuThrThrThrProAlaProProAspLysLeuGly 1085
3661 CGGTGCTTCGCTGACGAGTACGCTGCGCGG.....GGC 3627
1086 GlyLysGlnArgAlaAla.....PheLysSerGlyLy 1096
3626 CGCCACTCCCGCTCCGCGACGCGCTCTCCACCGCGCGCGCACCAAC 3577
1096 sArgValGlyLysProSerProLysAlaAlaSerSerProSerAsnProA 1113
3576 CCACCAAAAGCCCGCGAGACCGCGGCAAACTCACTCACCCTCC 3527
1113 laAlaLeuProValAlaSerAspSerProMetGlySerLysThrLys 1129
3526 CCGCCACACCGCGACACCTTACCAACCCATCGACCCCGCTCA... 3480
1130 GluThrAspSerProSerThrProGlyLysAspGlnArgSerMetIlele 1146
3479 TCACAGAGTCCGGCGCATCAACACCCCGCGGTCGCGTCCCTC 3430
1146 uArgSerArgThrLys.....ThrGlnGluPhe...H 1157
3429 CGCCCCCACCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3380
1157 isSerLysArgArgProSerGluGlyArgLeuProAsnCysArgAla 1173
3379 ACAGCATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3330

1174 ThrLysLysLeuLeuAspSerHisLeuProAlaThrPheLysValSe 1190
1329 AGCCGCGCCCGCCAGCCGACATCCAGCGCGCTGACCAACCCCAAGA 3280
1190 rSerSerProGlnLysGlnLysArgValSerGlnArgAlaArgValProL 1207
3279 CCGCACATCCCGCTGCGACGACGCCATCTCCCGACGCCCAAGGAGACGA 3230
1207 yProGlyAlaGlySerLysLeuSerAspArgProLeuHisAlaLeuLys 1223
3229 CACCGCGACAGCCTCA.....CCATCGACGCGCTCAAGC 3195
1224 Arg.....LysSerAlaPheMetAlaProValProHisLysLys 1236
3194 AGGTGGCGGCGCAACACTTCCTGCTCAACACCCGCTCCACCTCCAC 3145
1236 sarGlnLeuValLeuArgHisGlySerSerSerSerSerAlaSerA 1253
3144 CCGAACACATCGCGCGCCACACCGCTCCAAACCCCTCCGCGACACATCAA 3095
1253 La..... 1253
3094 CCGGCGCGCGCTCGCCCTCGACACACACACGCGCCGACAGCCCC 3045
1254MetGlyGluMetGlyArgArgArgLysLeu..ArgValPr 1266
3044 CCTACACACACCGCCCGAAAGGAGTCCACACCTCCGCTAGCGCACACC 2995
1266 oProProSerSerArgLysLys..LeuLeuProArgLysPro..... 1279
2994 GTCCACACACTTCAAGCTGCTGCTGCTACCGCGCCCATTCGCCAAGCCG 2945
1280SerProProArgAlaMetAla 1286
2944 ACAACACGACACACACGACGCGCCCTCATCCACCCCAAGACCATGCC 2895
1287 SerLeuProGlnSerSerHisPro.....ArgArgProProPh 1299
2894 ACCACCCCGCATCACACACTGACCGCGCTCCACAAAGGACACACCCA 2845
1299 eLeuProGlnAla..... 1303
2844 CATCCCGGAACCCCGGAACCTCCCGACATCCCTCCACATCCAGCC 2795
1304ArgLeuSerAlaAlaPheGlnGlyAlaMetLysThrLysVal 1317
2794 AAATCGCTCCGCTCAAAACGCAATAGCTGCGACGCTCCACCCGCAAT 2748
1318 LeuProProArgLysGlyArgGlyLeuLysLeuGlnAlaLevalGlnLys 1334
2747 ..CCATCACCGACCTCGCGAGCAGTCCCTCGAACACACCGGCCACATC 2701
1334 sileThrSerProSerLeuLysLysPheAla..... 1344
2700 AACCGCGCTCCGACGACGACAACTCCGCGAGGCGGTATGACCGACC 2651
1344 1344
2650 GTGCCTCCGCTGCTCGCGCGACGCGGAGATGCGCGCGCGCGGTGCA 2601
1345CysLysAlaProGlyAlaSerProGlyAsnProLeuSe 1357
2600 CTGACGAGATCTCTGCGAGGCGACGACGCTGCGTGGGCGCGATTC 2551
1357 r.....ProSerLeuS 1361
2550 GAGGAGGTGTGACGCCCTCTTCTGAACCATGAGATGCCGTCCGC.. 2503
1361 eAspLysAspArg..... 1365
2502 ..GAACGAGACGCTGCTGCGCGGTGTGACCCAGTATCCGGGAGTGA 2455
1366 ...GlyLeuLysGlyAlaGly.....GlySe 1373

2454 CATGCTCTCGCGGAGGCGCGCCCGGTGACGTTGAGACGACGCGAA 2405
1373 rProValGlyValGlnGlnLysLeu..... 1381
2404 TCCGCGGCGCTGAAGCTACTCTCTGCGCGCGCGGCGGAGAGTCCGCC 2355
1382ValAsnValGlyThrGlnLysLysLeu 1390
2354 AACATGGCTGCATGTGCGGCGAGTGAAGAGCGTGGCTGTCGACGCCG 2305
1391 ProThrSerGlyAlaAsp.....ProLeuLysArg..... 1400
2304 CCGGCTGCGCGCTCTGCTGCGCGCCATTCGTCGCGAGGTCCGACGACCG 2255
1401AsnProThrAsnArgSerL 1407
2254 CTTCCTGCTCCCGGAGAGACATGACCGCGCGCGCTTCACCGCGCG 2205
1407 eLysGlyLys.....LeuMet..AsnSerLysLysLeuSerSerTh 1420
2204 TGCGGACCGCGGATGCTATGCTGCGGACGCGGAGGTCTGCTCTC 2155
1420 rAspCysPheLysThrGlnAla..PheThrSerProGlnAlaLeuGlnPro 1436
2154 GAGCGCTCGATGCCACCATGCTCGCGCGGACGCGGACCTTCATCA 2105
1437 GlnGlyThrAlaLeuAlaProLysLysArgSerArgLysArgAlaG 1453
2104 GCGGCGCTCGTGGACACACAGTGCACCGCGCTCGGACGAGACAGACATC 2055
1453 Y...AlaHisGlyLeuSerLysGlyProLeuGlnLysArgProLysLeuG 1469
2054 CCGCGACATGGG.....CGCGCGCATTCACACGAGATGAGATGCCGAG 2011
1469 lProAlaLeuLeuLeuThrProArgAspArgLysSerGlyThr..... 1483
2010 GACGTATGCGGCGGTGACGACCCGAGTCTCCAGACCGGACGACGCCA 1961
1484GlnGlyAlaSerGluAspAsn.....SerG 1492
1960 CTCGAAAGCGAACAGCGCGCGCTGCGGGAACCGCTGCTCCATCAGC 1911
1492 yGlyGlyLysLysProLysMetGlnGlnLeuGlnLysProAlaSerGlnP 1509
1910 CGGCTCGGAGAGTCTCGCGTGCAGAGATGCTCCCGAGCGCAGG 1861
1509 rProGlnGlyArg.....ProCysGlnProGlnThrArgAlaGlnLys 1523
1860 GGCACCGGCGGTGCGCGCGCGCTGCGCTCCCGCGACATCTGTCGA 1811
1524 GluProGlyHis..ThrAsnLysSerSerLysSerLysArgLysArgLeu 1539
1810 TGCCCTGGGCGAAGCGGCTAGCGCTGATCGATCGCGGCCATGCC 1761
1540 ThrArgGly.....ArgAlaLysAsnThrThrSerSerPro.... 1551
1760 GCGCGTGGGTTCCTCGCGGCGAAGTACGCGCGATTCGCCCGAGGT 1711
1552CysLysGly..ArgAlaLysArgArgArgGln 1562
1710 GGTTCCTCCCTGACGACGCGCGCGCGCGCGCGCGCGCGCTGCT 1661
1563 GluValLeuProLeuAspProAlaGlnProGlnLysLeuLysLysTyr 1579
1660 CGATGCTGCGAGAACTGCTCGGCTCTCGGCGACAGACACCGCAC.. 1613
1579 eSerSerCysLysArgLeuArgSerAspSerArgThrProAlaPheSerP 1596
1612GATCTCGACACCGACCC...GCTC 1591
1596 rPheValArgValGlnLysArgAspAlaPheThrThrLysThrVal 1612

```
1590 CGACACCAAGCCGCCGCCGCCAGCG..... 1562
1613 ValAsnSerProGlyAspAlaProGlySerProHisArgLysProSerSerSe 1629
1561 ...GACTCACCACCCACCATCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 1515
1629 rAlaSerSerSerSerSerSerSerSerSerPheSerLeuAspAlaAlaGlyA 1646
1514 CGCAACCGGACGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 1487
1646 lAserLeuAlaThrLeuProGlyGlySerIleLeuGlnProArgProSer 1662
1486 ..... 1475
1663 LeuProLeuSerSerThrMetHisLeuGlyProValValSerLysAlaLe 1679
1474 GCGCGGACACCCACCCAG.....GCACACCC..... 1448
1679 uSerThrSerCysLeuValCysCysLeuCysGlnAsnProAlaAsnPheL 1696
1447 .....CGACACCGACACCCACC 1429
1696 ysAspLeuGlyAspLeuCysGlyProTyrTyrProGluHisCysLeuPro 1712
1428 CGACCCCACTCCGCGAGCGCGCGCAC.....CCCCACCGCGGCC 1385
1713 LysLysLysProLysLeuLysGlyValArgProGluGlyThrCysGL 1729
1384 CGACGCGCTCCT.....CCAAATCACATGCGCA 1356
1729 uGluAlaSerLeuProLeuGluArgThrLeuLysGlyProGluCysAlaA 1746
1355 TTGTCGCCACTCACCC.....CGAACGACAGACCCGCCGCCGCCGCCGCC 1312
1746 la.....AlaAlaThrAlaGlyLysProPro 1754
1311 CGGACCCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1274

seq_name: /SDS2/cgdata/geneseq/geneseq/VAl1997.DAT:AAT93095
seq_documentation_block:
ID AAT93095 standard; cDNA; 24379 BP.
AC AAT93095;
XX
XX 11-MAY-1998 (first entry)
XX Streptomycetes frenolicin gene cluster.
XX Frenolicin; antibiotic; feed additive; antitoxicidial;
KW coccidiostatic; efflux pump; butyrate starter synthase;
KW polyketide synthase; PKS; hemiketalase; ketoreductase; cyclase;
KW dehydrase; ketoreductase; hydroxylase; Streptomycetes roseofulvus;
KW ds.
XX Streptomycetes sp.
OS
XX Key Location/Qualifiers
FH 636..2948
CDS /tag= a
FT /product= 80 kDa non-membrane protein
FT /note= "gene A (specifically claimed)"
FT 2945..3916
CDS /tag= b
FT /product= membrane protein
FT /note= "gene B (specifically claimed)"
FT 4020..4844
CDS /tag= c
FT /product= protein with 6 membrane-spanning domains
FT /note= "gene C (specifically claimed)"
FT 4841..6415
CDS /tag= d
FT /product= ATP-binding component of ABC transporter
```

```
FT /note= "gene D (specifically claimed)"
CDS 6533..7183
FT /tag= e
FT /product= unknown non-membrane protein
FT /note= "gene E (specifically claimed)"
CDS 7344..8897
FT /tag= f
FT /product= putative efflux pump
FT /note= "gene F (specifically claimed)"
CDS 9164..10012
FT /tag= g
FT /product= transcription activator
FT /note= "gene G (specifically claimed)"
CDS complement (10105..10621)
FT /tag= h
FT /product= translationally coupled to gene I
FT /note= "gene H (specifically claimed)"
CDS complement (10618..11628)
FT /tag= i
FT /product= homologue of fabH
FT /note= "gene I (specifically claimed)"
CDS 11809..12066
FT /tag= j
FT /product= acyl carrier protein
FT /note= "gene J (specifically claimed)"
CDS complement (12154..13209)
FT /tag= k
FT /product= acyltransferase
FT /note= "gene K (specifically claimed)"
CDS 13409..14686
FT /tag= l
FT /product= PKS ketoacylsynthase subunit
FT /note= "gene L (specifically claimed)"
CDS 14767..16047
FT /tag= m
FT /product= PKS chain length factor
FT /note= "gene M (specifically claimed)"
CDS 16120..16371
FT /tag= n
FT /product= acyl carrier protein used by the PKS
FT /note= "gene N (specifically claimed)"
CDS complement (16453..16935)
FT /tag= o
FT /product= putative hemiketal dehydrase
FT /note= "gene O (specifically claimed)"
CDS 17088..17903
FT /tag= p
FT /product= ketoreductase related to actIII
FT /note= "gene P (specifically claimed)"
CDS 17903..18898
FT /tag= q
FT /product= cyclase/dehydrase related to act VII
FT /note= "gene Q (specifically claimed)"
CDS 18895..19839
FT /tag= r
FT /product= cyclase/dehydrase related to activ
FT /note= "gene R (specifically claimed)"
CDS complement (19950..20907)
FT /tag= s
FT /product= oxidoreductase
FT /note= "gene S (specifically claimed)"
CDS complement (20904..22094)
FT /tag= t
FT /product= quinone-forming hydroxylase
FT /note= "gene T (specifically claimed)"
CDS complement (22505..22179)
FT /tag= u
FT /product= not specified
FT /note= "gene U"
```

EP806480-A2.
12-NOV-1997.

This DNA sequence comprises the streptomycetes *freemollicus* gene cluster containing specifically claimed coding sequences (genes A-V) that respectively encode 21 proteins (see AM341199-219) involved in *freemollicus* synthesis. The genes can be divided into 5 subclusters: (1) genes A, B, C, D and F encode an efflux pump; (2) genes H, I, J and K encode butyrate starter synthases; (3) genes L, M and N encode polyketide synthases (PKS); (4) genes O, P, Q and R encode a hemiketalase, a ketoreductase and cyclase/glycosylases; and (5) genes S and T encode a keto/enoyl reductase and a hydrolase. Also claimed are vectors, host cells (especially a *Streptomyces* sp., particularly *Streptomyces roseofulvus*), and the encoded proteins. Cells transformed using the above sequence can be cultured to produce *freemollicins* or *freemollicin* precursors. The precursors can be converted to *freemollicins* by chemical or other methods. The *freemollicins* can be oxidised to *freemollicin B*, an antibiotic used as an antitubercular agent. The *freemollicins* can be used as animal feed additives.

alignment_block;
US-09-508-821B-6 x AAT93095

Align seg 1/1 to: AAT93095 from: 1 to: 24379

```

22  GInGInIhtrSerArgLeuGlnAsnTyrArgIn.....Pose 34
      |||||:::|||||
      CAGAAAGCTGGCTCAGCAGACAGTACAGCAGCGCGAATGGCCAGACCTTCC 11884
11835 CAGAAAGCTGGCTCAGCAGACAGTACAGCAGCGCGAATGGCCAGACCTTCC
      |||||:::|||||
34  rGlnAlaGlyLeuSerCysAspArgGlnArg.....LeuDeuAlaL 48
      |||||:::|||||
      ACGGGAGAGGCCCTCGACACAGACTTCAACCACTCGGGTTAGAGACTCGCG 11934
11885 ACGGGAGAGGCCCTCGACACAGACTTCAACCACTCGGGTTAGAGACTCGCG
      |||||:::|||||
48  yAspTyrTyrAsnProGlnProTyrProSer.....58
      |||||:::|||||
11935 ACGGTCTACAGAGnTCGnCAACCCGATCCAGAGACAGACAGnATCACCgt 11984
      |||||:::|||||
59  TyrGInGlyGlyAlaGlyIhtrProSerGlyIhtrAlaAlaAlaValAlaAla 75
      |||||:::|||||
11985 GCCGACGACGAGAGTGGnACTCTCGACACCCCGCGCCCTCATCGCGT 12034
      |||||:::|||||
75  aAspLyIhtrIhAspArgIySerIyAlaLeuProIhtrGInGInIyLeug 92
      |||||:::|||||
      ACGTCGACGCGCGCGCGAGGTTCCCGCACCCTACCCGACGGAGAGGT...G 12081
12035 ACGTCGACGCGCGCGCGAGGTTCCCGCACCCTACCCGACGGAGAGGT
      |||||:::|||||
92  InGlyArg.....ProAlaPheProGlyIhtrGlyAlaGlnAspSer 106
      |||||:::|||||
12082 CGGGAGAGGGGGGGGCGCGAGACGGGCGGGGCGCGCGCGGTACGACGCGG 12131
      |||||:::|||||
107  ProTyrProGlnIhtrTyrAla.....113

```

12969 ACGGCTGGCGACAGCGGCGCTTCCAGCGCTCTCTCGCGCGCGCGCG 13018
358 SerThr.....ProSerProLeuMetPro..... 365
13019 TCCAGCCACCGCGCGCGCACTCTCCGCCCGAGTGCACACACGCGGAA 13068
366AsnLeuGluAsnProPheProThrSerGlnG 375
13069 GCATCTCTCCACCGCGGTGGTGAAGTCCGCGCACCGCGGTGGAGTCCG 13118
375 InProLeu.....Ser 378
13119 CGCCATCCCGCGCGCTCGCGCGCTTACCGCGAGCAGCGGCGCAC 13168
379 ThrGlyAlaPheProAlaGlyLeuThrAspHisSerHisPheMetPro.L 395
13169 ACGGGGCGCGCTCTCTCTCGTACCGCTCGCTCGACACTTTCGCGC 13218
395 euLeuAsnProSerProThrAspAlaThrSerSerValAspThrGlnAla 411
13219 TCCTGCTCCG.....CGACTCTCTGCC.....GGTACCGAGCCT 13253
412 GlyAsnCysLys.....ProLeuGlnLysAspLysLe 422
13254 CGCCCTTCGCGCTCGACGCGGTGGACTCCCGCTCCAGCGCGC..... 13298
422 uProGluAsnLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuT 439
13299 .CGCGCGCGCGCGCGCGCTCAAGCGCTTCAAGCGCGCTTCAAGCGCG 13347
439 hrLeuGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeuSer 455
13348 AGGTCGGGTGGAGCGGTGAACGGCAC..... 13376
456 LysAlaAlaValProGlnLysGlyValLysAsnLeuValSerArgTh 472
13377 ...GCGCGCTCCAGCAGCGCGCAAGGGGTGCTGAGTGAACCGACAGTC 13423
472 rProGluGlnHisLysSerGlnHisCysSerPro..... 483
13424 GCGG.....TCACCGGCATCGCGGTGGTGGC 13449
484GlutylserGlyTyrSerAlaGluProAlaGlyThrProLeuSer 498
13450 TCGGGAGGATCGCGCGCAAGCGTACTGGAGCAGCTACCTCCGCGAC 13499
499 GluProProSerThrProGlnSerThrHisAlaGluProGlnGluAl 515
13500 GCACCGCCACCGCGCGCTCTCTCTTTCGACGCTCGCGCT..... 13541
515 aaSerThrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrC 532
13541 13541
532 ysAsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLys... 547
13542TCGCTTCGCGATCGCGCGAGTTCGACTTCGACCGCGCGCG 13586
548AlaLysProGluSerValSer..... 554
13587 CGGGCTGAGCGCGCGAGTTCGCGCATGAGCGCGCGCGCGCGAGTTC 13636
555ThrCysSerValThrSerP 561
13637 GCGGTGGTGGCGCGAGGAGCGCTCGCGAGCAGCGGACTCGACGTGC 13686
561 roAspAspMetSerThrLysSerAspAspSerPheGlnSerLeuHisG 577
13687 CGACTTCACCGCGCGATCGGGGTGGGTCATCGCGAGCGCGGTGGCG 13736
577 lysLeuProLeuAspSerPheSerLysPheValAla..... 589
13737 GGACCACTCGCTGGAA.....CGGAGTACCTCGCTTCAGCGACAGC 13780

590GlyGluArgAspCysProArgLeu.....LeuLeuSe 600
13781 GGGGACAGTGGGAACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13830
600 rAlaLeuAlaGlnGluAspLeuAlaSerGlnLeuLeuGlnGluAla 617
13831 GCGCTTCACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13873
617 laileGlyGluLysAlaAspLysAlaTrpAlaGluAlaProSerLeuVal 633
13874 ...ATCGGC.....GCGAGGCGCGCGCGCGCGCG 13900
634 LysAspSerSerLysProProPheSerLeuGluAsnHisSerAlaCysL 650
13901 GTCTCCACCGGTGCACCTTCGCGATCG..... 13928
650 euAspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAla 666
13928ACTCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13967
667 LeuProAspSerLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSe 683
13968 GCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 13982
683 rProGlyLeuPheGluAspProSerValAlaPheAlaThrProAspProL 700
13983TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14010
700 ysLysThrThrGlyProLeuSerPheGlyThrLysProThrLeuGlyVal 716
14011 GCGGATCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14048
717 ProAlaProAspProThrThrAlaAlaPheAspCysPheProAspThrTh 733
14049 COTCG 14098
733 rAlaAlaSerSerAlaAspSer..... 740
14099 GAACGCAACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14148
740 740
14149 GAGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14198
741 ...AlaAsnProPheAla..... 745
14199 CGGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14248
746TTPProGluGluAsnLeuGlyAspAlaCysProArgT 758
14249 CACGGCAGGAGATGGCGAGGCGATCC...GGCACGGCTCGCGAGAG 14295
758 rGlyLeuHisProGlyGluLeuThrLysGlyLeuGlnGlnGlyLys 774
14296 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14335
775 AlaSerAspGlyLysSerLysGlyAspThrHisGluAlaSerAlaCysLe 791
14336 ...GECTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14380
791 uGlyPheGlnGluAspProProGlyGluLysValAlaSerLeuProG 808
14381AAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14416
808 lysAspPheLysGlnGluValGlyGlyValLysGluGluAlaGly... 823
14417GTCAGCTCCATCAAGTATGATCGCGCGCGCGCGCG 14446
824GlyLeuLeuGlnCysProGluValAlaLysAlaAspArgTrpLe 838
14447 TCGCTCGCTCCATCGGTCTGCTGGAGATCGCG.....GCTCGG... 14485

15964 ACGGCGTCTGCTCTCGCGCGGCCACGGCGCTTCAACGGCGCGCTCT 16013
 1314 LysThrLysValLeuPro.ProArgLysGlyArgGlyLeuLysLeuGluA 1330
 16014 CGTCCGCGGCGCGCGCGCGCGCTTACCGCTCGAGG... 16050
 1330 LaileValGlnLysIleThrSerProSerLeuLysLysPheAlaCysLys 1346
 16051CCT 16053
 1347 AlaProGlyAlaSerProGlyAsnProLeuSerProSerLeuSerAspLys 1363
 16054 GCCCGGCTGCGCGCGG...CGCGCGCACCCCGCTCCACC... 16092
 1363 sAspArgGlyLeuLysGlyAlaGlyLysSerProValGlyAluGluLug 1380
 16093GCGCGCGCACCCCGAGAGTTCCTATGA 16123
 1380 LysLeuValAsnValGlyThrGlyLysLeu.....ProThrSerGly 1394
 16124 GCGACTGACGCTGACGACCTCAAGAACTGCTCCGCGAGACCGCGGG 16173
 1395 AlaAspProLeuCysArgAsnPro.....ThrAsnArgSerLeuLys 1408
 16174 GAGGACGACAG...CGTGCACCTGCGCGGAGAACTCGACAGCCCTCG 16219
 1408 sGlyLysLeuMetAsnSerLysLeuSerSerThrAspCysPheLysT 1425
 16220 TGGACCTCGGTACGACTCCCTCGCGCTCTGGA..... 16253
 1425 hrGluAlaPheThrSerProGluAlaLeu.....Gln 1435
 16254 ..GACGGCGCGCTGCTCCAGAGCGGTACGGCATCGCGTACTGACGA 16301
 1436 ProGlyGlyThrAlaLeuAlaProLysLysArgSerArgLysGlyArg... 1451
 16302 GACGCTGCGCGCTGGCGGCCCGCGCGGAACTGCTCGAGAGGTCAACA 16351
 1452 AlaGlyAlaHisGlyLeuSerLysGlyProLeuGluLysArgProTyrL 1468
 16352 CCACCGCGGCCACCGCTGAGC..... 16373
 1468 euGlyProAlaLeuLeuLeuThrProArgAspArgAlaSerGlyThrGln 1484
 16374GGCGGCGCACGCGAAGGAAG 16394
 1485 GlyAlaSerGluAsnSerGlyGlyGlyLysLysProLysMet... 1500
 16395 GGTCCGCGCCACCCCGCGCGGCGGTGGGCGGACCTTCGTTCGC 16444
 1501GluGluLeuGlyProAlaSer..... 1507
 16445 GCGGACCTCACAGAGGTGCGTCTGGGTCACTCGCGCGACAGCGT 16494
 1508GlnProGluGlyArgProCysGln..... 1516
 16495 GCGAGTTCCTCCGCGCTCCGACACGAGCTCCCGTACGACGAACCTGG 16544
 1516 1516
 16545 GACGATCACGGGAGCGCGGTCTCGCGCGGTCAACACGACGCGCG 16594
 1517ProGluThrArgAlaGlnLysGlnProGlyHisThrAsnTyrS 1531
 16595 TAGAACTCGCTCGACGCGCGCTCTCCCGGCGGTACGTCCAGCAT 16644
 1531 erSerTyrSer.....LysArgLysArgLeuThrArgGly... 1542
 16645 CGAGAACCAAGTCCGCGCTGTGACGCGGTGCTCGGGAACCGCGGCGGT 16694
 1543ArgAlaLysAsnThrThrSerSerPro..... 1551
 16695 TGAAGCGCGCACCTCTCCGCGATCGCGCGCTCTCGCGCGCGCGCG 16744

1552 .CysLysGlyArgAlaLysArgArg.....ArgG 1561
 16745 GTGTGGGGGAGTGGCGAAGAGCCCTCCCGCGCAAGCGCGCGGAA 16794
 1561 LncLncLncValLeuProLeuAspProAlaGluProGluLysLeuLys 1577
 16795 CTCTCGCGCGGTACCGCTCCAGGTGGTGGTGGCTGGCGGTAGAACG 16844
 1578 TyrIleSerSerCysLysArgLeuArgSerAspSerArgThrProAlaPh 1594
 16845 CCGTACTCTCGGTAGAGCGCGCGGAACGCGCGCTCGCGCGCGAGG 16894
 1594 eSerProPheValArgValGluLysArgAspAlaPhe..... 1606
 16895 GACC.....CGGCCCTCAGCTCGTCCGCGCGGTGGTGGTGGTGCAC 16935
 1607ThrThrIleCysThrValValAsnSerProGly..... 1617
 16936 GAGAGCTCCAGAGGTGT.....CTGACCTGTCCGCGCAAGCTCCG 16979
 1617 1617
 16980 CGCGTGGTCTGACGGCGCTTGAAGCCCTCAAGGAGCGCGCGGAGG 17029
 1618AspAlaProLysProHisArgL 1625
 17030 CTGCAACACGACCGCGCTGACCGACCGCGCGCGCGCGCGAGGA 17079
 1625 ysProSerSerSerAlaSerSerSerSerSerSerSerSerPheSerLeu 1641
 17080 AGCCATCATGACACAC..... 17096
 1642 AspAlaAlaGlyAlaSerLeuAlaThrLeuProGlyGlySerIleLeuGl 1658
 17097GCAGCTCCGACACCGCGCGCGCGCGCGCGCGCGCGCG 17134
 1658 nProArgProSerLeuProLeuSerSerThrMethIstLeuGlyProValV 1675
 17135 CGCGGAGCGCGCTCTGTCGACCGCGCGCGCGCGCGCGCGCGCG 17184
 1675 alSerLysAlaLeuSerThrSerCysLeuValCysCysLeuCysGlnAsn 1691
 17185 TCCTCGAGGCGCTCTGTCGCGCGCGCTATCCCGTGGTGGTCTGCGCGCG 17234
 1692 ProAla 1693
 17235 GAGCGC 17240
 seq_name: /SIDS2/gcgdata/geneseq/NA1998.DAT:AAV25925
 seq_documentation_block:
 ID AAV25925 standard; cDNA; 24379 BP.
 XX
 AC AAV25925;
 XX
 DT 15-JUL-1998 (first entry)
 XX
 DE Streptomyces roseofulvus frenolicin gene cluster.
 XX
 KW Streptomyces roseofulvus; frenolicin gene cluster; frenolicin B;
 antibiotic; ss.
 XX
 KW Streptomyces roseofulvus.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 636..2948
 FT /tag= a
 FT /trans_except= (pos:636..638,aa:Met)
 FT /note= "encodes protein given in AAV55800"
 FT 2945..3916
 FT CDS /tag= b
 FT /trans_except= (pos:2945..2947,aa:Met)
 FT

12232 GGGAGCGCGGCGAGAGCCGAACACACAGCGCGCGCGCGCGAGCGAGCG 12281
133 ProLeuProAlaGlyValAlaGlyTyrAspGluAsnLeuMetLysLysTh 149
12282 CCGGATACGCGCGCGCGCGCGCGCGAGACTCGCTCCGCGGAGCGGTG 12331
149 rAlaValPro.....ProSerArgGlnT 157
12332 AGCAGCACCGACCGACCGCGCGCGCGCGCGTGTGCTCGCGCGTCCGCGTCC 12381
157 yAlaGluGlnGlnValAlaGlnValProPhe.....Arg 167
12382 GTTCGCGAGCAGGTTCCTCCAGCGCGCGCGCGAGACAGCGCGCGCGCG 12431
168 ThrHisSerLeuHisValGlnGlnProProProProGlnGlnProLeuAl 184
12432 GCTGACCGCACCGACAGCGCGCGCGCGCGCGTCCGCGCGCGTGTGAGCGCG 12481
184 aTyProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerProL 201
12482 CGACCGGTCCG 12531
201 euProPheProGlnGlnThrHisPheProGlnHisSerGlnSerPhePro 217
12532 GACCGGTTCGCGCGTGG.....CCCTGGTTCAGACGCGCGCGCGCGCG 12572
218 ThrSerSerThrTySerSerSerValGlnGlnGlyGlnGlnGlnAlaH 234
12573 CGCGAGCGCGCGGAGTGAAGAGTACGTCCCGCTGAGCGCGCGCGCGCG 12622
234 sSerTyLys.....SerCysThrAlaProT 243
12623 GTACGCGCGTGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 12672
243 hrAlaGln.....ProHisSerArgProLeu.....ThrAlaSerSer 256
12673 CGCGCGCGACAGCACGCTGTGACGGCGCGCGCGCGCGCGCGCGCGCGCG 12722
256 rLeuAlaProGlnArgValGlnAsnLeuHisAlaTyrgLysGlyA 273
12723 CGTCG 12772
273 rGlnSerTyAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 289
12773 AGCATCCCCCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12822
290 GlnGlnAlaLeuGlnSerArgHisAlaGlnGlnGlnGlnGlnGlnGln 306
12823 CAGCGGAGCG 12846
306 nasLeuAlaLysTyrgLysHisTyrgLysGlnGlnGlnGlnGlnGlnGln 323
12847GCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12873
323 lnProAspAlaAlaValArgThrProGlnGlnTyTyr.....GlnThrPhe 338
12874GAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12918
339 SerProSerSerHisSerPro..... 346
12919 CG 12968
347AlaArgSerValGlyArgSerProSerTySer 357
12969 ACGGCTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13018
358 SerThr.....ProSerProLeuMetPro..... 365
13019 TCCAGCG 13068
366AsnLeuGluAsnPheProTySerGlnG 375

13069 GCATCTCTCCACCGCGCGTGTGAACCTCGCGCACCGCGCGTGTGAGTCCGG 13118
375 lnProLeu.....Ser 378
13119 CGCCATCCCGGCGCGCGTGTGCGCGCGCTGACCGGCGACACAGCGCGCAC 13168
379 ThrGlyAlaPheProAlaGlyIleThrAspHisSerHisPheMetProL 395
13169 ACCGCGCGCGCTCTCTCTCGTACCGCGCGCTGCTCGCGACTTCTCGCGC 13218
395 euLeuAsnProSerProThrAspAlaThrSerSerValAspThrGlnAla 411
13219 TCTCTGTCGCG.....CGACTCTCTGCC.....GCTACCGAGCCT 13253
412 GlyAsnCysLys.....ProLeuGlnLysAspLysLe 422
13254 CGCGCTTTCGCGTTCGACGCGCGGTGTGACTCCGCTCCAGCGCGCGC..... 13298
422 uProGluAsnLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuT 439
13298 CGCGCGCGCGCGCGCGCGCGCTGACCGCGCTGACCGCGCTTCCGAGGAGCTTCC 13347
439 hrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeuLeuSer 455
13348 AAGTCTGGGTTCGAGGCGCGTGAAGCGCAC..... 13376
456 LysAlaAlaValProGlnLysGlyValLysAsnLeuValSerArgTh 472
13377 ...CGCGCGTTCAGGACCGCGCGCGGTGTGACGATGACACACAGTC 13423
472 rProGluGlnHisLysSerGlnHisCysSerPro..... 483
13424 GCG.....TCACCGCATCGCGGTGTGTGGC 13449
484GluGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSer 498
13450 TCGGAGGAGTGTGGCGCAAGCGTACGAGGAGGAGTCACTCCGCGGAC 13499
499 GluProProSerSerThrProGlnSerThrHisAlaGluProGlnGlnAl 515
13500 GCACGCGCGCGCGCGCTCTCTCTTCTGAGCGCTCGCGCT..... 13541
515 aAspTyLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrc 532
13541 13541
532 ysAsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLys... 547
13542TCCGCTCGCGGATCGCGCGCGGAGTGTGACTTCGACCGCGCGCG 13586
548AlaLysProGluSerValSer..... 554
13587 CGGCGCTGTGAGCGCGCGGAGTGTGCGCGCATGGACCGCGCGCGCGAGTTC 13636
555ThrCysSerValThrSerP 561
13637 GCGTGTGTGAGCGCGGAGGAGCGCTCGCGGAGCGGAGTGTGAGTGTGCG 13686
561 rGAspAspMetSerThrLysSerAspAspSerPheGlnSerLeuHisG 577
13687 CGACTCTGCGCGCGCGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 13736
577 LysSerLeuProLeuAspSerPheSerLysPheValAla..... 589
13737 GAGCACCTCGTGGAA.....CGGAGTACTCTGCGCTCAGCGAGCAGC 13780
590GlyGluArgAspCysProArgLeu.....LeuLeuSe 600
13781 GGGGAGCAGTGGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 13830
600 rAlaLeuAlaGlnGluAspLeuAlaSerGluLeuLeuGlyLeuGlnGluA 617
13831 CGCTTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13873

[illegible]

10

06-JAN-1999; 99US-0115435.
05-FEB-1999; 99US-0118848.
05-JAN-2000; 2000US-047962.
(REGC) UNIV CALIFORNIA.
Shen B, Du L, Sanchez C, Chen M, Edwards DJ;
WPI; 2000-465974/40.
P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
AAB07576, AAB07577, AAB07578.
New bleomycin gene cluster components useful for peptide and/or
polyketide metabolites, especially bleomycin, production and for
chemically modifying biological molecules -
Claim 8; Page 97-136; 162pp; English.
The present sequence represents the BLM (Bleomycin) gene cluster,
containing open reading frames (ORFs) 8-30. The proteins encoded
by the gene cluster are useful for producing peptides and/or polyketide
metabolites, especially bleomycin or bleomycin analogues. They are
also useful for chemically modifying biological molecules to produce
branched methyl groups, and for coupling amino acids and fatty
acids. They may be reacted with an apo-carrier protein and coenzyme A
to produce a holo-carrier protein. The BLM gene cluster or catalytic
domains can be used individually or collectively to produce
thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
microbial metabolites. The BLM gene cluster may also be used to produce
sugars.

50857 BB. 7256 A. 25139 C. 19353 G. 7109 T. 0 other;

```

lignment_scores:
  Quality: 468.00
  Ratio: 0.559
  Percent Identity: 21.368
  Length: 2120
  Gaps: 103

```

ulignment_block:

Align seg 1/1 to: AAA58471 from: 1 to: 58857

```

32  GlnProSerGlnhlaGlyLeuSerCysAspArgGlnArgLeuLeuAlaLay 48
      ::::::::::: :::: :::::::::::::: ::::
51249  CGACCTCGGGAGCCCGCTCCGCCGACGCGACCGCTCTGGCTGGCGC 51298
      ::::::::::: :::: :::::::::::::: ::::
48  sAspTyrTyrAsnProGlnProTyrProSerTyrGluGlyGlyAlaGlyT 65
      ::::::::::: :::: :::::::::::::: ::::
51299  TGCTCGCGGCGGACGAGGAGCA...CCCTCGGCGCGG.....A 51333
      ::::::::::: :::: :::::::::::::: ::::
65  hProSerGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArgGly 81
      ::::::::::: :::: :::::::::::::: ::::
51334  CGCCACCGGGCGCGCTCCACCTGTGCGGCTCCACGACGACGACGGT 51383
      ::::::::::: :::: :::::::::::::: ::::
82  SerLysAla.....LeuProThrG1 88
      ::::::::::: :::: :::::::::::::: ::::
51384  GCTCGGCTACGCGCCACACCTCTCTCTCGACACCGCTCTCTACG 51433
      ::::::::::: :::: :::::::::::::: ::::
88  nGlnGlyLeuGlnGlyArgProAlaPheProGlyTyrGlyValGlnAspS 105
      ::::::::::: :::: :::::::::::::: ::::
51434  CCGTGTCTGGCGGGGTGTGCCA.....GGCGTAGC..... 51464
      ::::::::::: :::: :::::::::::::: ::::
105  erSerProTyrProGlyArgGlyTyrAlaGlyGluGlnSerLeuGlnAlaTrp 121
      ::::::::::: :::: :::::::::::::: ::::
51465  .CGGCGCGTTCGCGCCGCGGAGCACTACCGCGACGCGCACACCGCTGCCGA 51512
      ::::::::::: :::: :::::::::::::: ::::
122  GlyAlaProGlnProPro.....LeuProThrG1 127
      ::::::::::: :::: :::::::::::::: ::::

```


1350 a..... 1350
55493 ACCGCCGAGAGTGTGGGCTCTCGAAGAGCCCTTCATGCTGGGGCC 55542
1350 1350
55543 ACCCCGTCCTCAAGCCGAGCGTCCCGCATCAACGCTGACGGGAC 55592
1351 SerProGlyAsnProLeuSerProSerLe 1360
55593 GTCCAGAGATCCAGTGGTCAACCCGAGACACCCCGCTTCCACAGCC 55642
1360 userAspLysAspArgGlyLeuLysGlyValAlaGlySerProValGly 1377
55643 TCATCCACGC.....CTTCAAGACCGTACGGGGATCCCAAGGTG 55683
1377 aGlu.....GluGlyLeuValAsnValGlyThrGlyGln 1388
55684 CTCACACACAGCTCAACACCAAGGAGCGGATCGGGAGACACCCGA 55733
1389 LysLeuProThrSerGlyAlaAspProLeuGlyAsnPro..... 1402
55734 GAGCGCCCTGCGACCTGCTCGGCTCCGGCTCGACACCACTGCTGCTCC 55783
1403ThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLys 1416
55784 CGGGCTCATCTGTCACGCGCGGACGCGCGCTCATGAGCCGCCCGC 55833
1416 ys.LeuSerSerThrAspCysPheLysThrGluAlaPheThrSerProG 1432
55834 GGGGCGACGCGACCGCGCGCGCGCTCGACGCG.....AACTGCGCGCG 55880
1432 uAla.....LeuGlnProGlyLysThrAlaLeuAlaProLysLysArg 1447
55881 ATCTGGGCGCGAGACCTCGCGAGAGAGGTGCGCGCGACAGGAGCT 55930
1447 exArgLysGlyArgAlaGlyAlaHisGlyLeuSerLysGlyProLeu... 1462
55931 CGCGCGCTGGGCGGCACTCCATCCAGCGCATCAAGATCAACCAACCGG 55980
1463GlnLys 1464
55981 TGGAGAGACTGTCGAGCGGAGCTGTCATCGGCTCTGTCAGAGCG 56030
1464 sarGProLysLeuGlyProAlaLeuLeuThrProArgAspArgAla 1481
56031 CGCACGCTGGCGCGCATGAGACCAAGCTCCAGCGCATCAAGGAGGA 56080
1481 erGlyThrGlnGlyAlaSerGlnAspAsnSerGlyGlyGlyLys 1497
56081 GCGGAGAC.....GCTGACACCGGA 56100
1498 ProLysMetGlnGluLeuGlyProAlaSerGlnProGlnGlyArgPr 1514
56101 CCT.....GCCCGGCTGCTGA...CCGATCGCGCGCC 56132
1514 ocYsGlnProGlnThrArgAlaGlnLysGlnProGlyHisThrAsnLys 1531
56133 TGGCGCTCTGCTCATGCG...CGAGCTCATCTCGACACTACGCTGG 56179
1531 exSerTyr..... 1533
56180 GAGACCACTCGGGCTGTGCGCGAGATCCCGCTCCCTGACCCCT 56229
1534SerLysArgLysArgLeuTh 1540
56230 GACCTCGGTGCGCCACAGTGGCGCGCGCGCCACAGTGGCGGTGAAC 56279
1540 rArgGlyArgAlaLysAsnThr..... 1547
56280 TCCGGGCGCTCGGCGCGACCGGTGCTCTCTCGCGAGGAGTGAAGAC 56329

1548ThrSerSerProCysLysGlyArgAlaLys 1557
56330 CGCGCGCGCGCGCGCGCTGCGCGAGAGCCCTTCCGCGCGGAGCGAGAC 56379
1558 ArgArgArgGlnGlnGlnValLeuPro..... 1566
56380 CGCGGAGCTCTTCTGACAGCCCGCGCGAGACGCTCAACCAAGCCCGG 56429
1567LeuAspProAlaGluProGlnLysArgLeuLysTyr 1579
56430 TATGAGCGAGCGAGAGATGCTGCTCGCT..... 56460
1579 LeuSerSerCysLysArgLeuArgSerAspSerArgThrProAlaPhe... 1594
56461CGACGAGCGCGGAGAACCGCTTGGCCGT 56490
1595SerProPheVal 1598
56491 GAGGACGAGACACGAGACCGGCTGCTGAAAGGCGCGCGCTGCTGC 56540
1598 ArgValGlnLysArgAspAlaPheThrThrLysThrValAlaAsn 1615
56541 CGCGCTGCGAGCGCGCTGATGCTCGG...ACTAGGAGTACGGCGTGGGA 56589
1615 exProGlyAspAlaPro...LysProHisArgLysProSerSerSerAla 1630
56590 GCGCGACACCGCGCGCGCTGCGCGACACCGGAGACTGCGCGCTCA 56639
1631 SerSerSerSerSerSerSerPheSerLeuAspAlaGlyAlaSer 1647
56640 CCTGGTCTGCTGACT..... 56654
1647 rLeuAlaThrLeuProGlyLysThrLeuGlnProArgProSerLeuP 1664
56655CGCGCGCGCGCGCGCGCTTCAACCGCG... 56681
1664 rLeuSerSerThrMetHisLeuGlyProValValSerLysAlaLeuSer 1680
56682TGGGCGCGAGCGCGCTGCAAAACCAACCAACGAG 56711
1681 ThrSerCysLeuValCysCysLeuGlyGlnAsnProAlaAsn 1694
56712 CGGAGG.....CGCTGCGCTGCTGAGCGCGCGGAGAC 56744

OW of: US-09-508-821b-6 to: Issued Patents_NA:* out_format: pfs
Date: Dec 4, 2001 2:01 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODE=frame+ptn_model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09508821/runat_04122001_114530_10116/app_query.fasta.1.1842
-B=Issued Patents_NA -OPM=fastap -SUFFIX=ptn -GAPOP=12.000
-GAPEXT=4.500 -MINMATCH=0.100 -LOCPCI=0.000 -LOCPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=ptc
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09508821@cgn1_1.111 -NCPV=6 -ICPV=3 -LONCLIG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -NO_XLPYX -WAIT -THREADS=1

Search information block:
Query: US-09-508-821b-6
Query length: 1755
Database: Issued Patents_NA*
Database sequences: 351203
Database length: 113238999
Search time (sec): 68.580000

Sequence	Strid	Orig	Zscore	EScore	Len	Documentation	1.6e-13	8438
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-945-283-1					508.00	379.92	1.6e-13	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-258-261B-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-456-837-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-642-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-642A-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-458-076A-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-335A-4					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-335A-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-729-214-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-028-934-6					493.00	358.47	2.6e-12	28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-764-233A-1					486.00	351.90	5.9e-12	49377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-144-085-3					466.00	332.28	7.4e-11	68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-335-409-1					462.50	333.63	6.2e-11	18782
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-443-501A-2					454.50	327.67	1.3e-10	71989
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-105-537-1					450.50	322.30	2.6e-10	44377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-227C-7					450.50	322.30	2.6e-10	13842
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-105-537-5					445.50	319.96	3.6e-10	36778
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-320-878-19					445.50	319.96	3.6e-10	36506
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-227C-13					442.50	325.40	1.8e-10	13987
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-428-517-1					429.00	309.26	1.4e-09	50937
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-227C-7					426.50	303.81	2.8e-09	44377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-198-1					426.50	303.81	2.8e-09	44377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-370-700-1					422.50	295.99	7.7e-09	80161
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-227C-1					419.50	301.56	3.8e-09	30001
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-474-933-1					415.00	301.25	3.9e-09	20235
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-439-009A-3					415.00	301.25	3.9e-09	20235
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-105-537-5					411.50	296.46	7.3e-09	15872
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-320-878-19					409.00	291.47	1.4e-08	38506
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-227C-1					400.00	283.60	3.8e-08	44377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-105-537-1					399.50	245.36	5.1e-08	441529
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-103-840A-1					397.50	244.61	3.3e-08	30001
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-474-933-1					397.50	244.61	3.3e-08	30001
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-458-568A-11					394.00	289.26	1.8e-08	12001

seq_name: /cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-945-283-1

seq_documentation_block:
Sequence 1, Application US/07945283
Patent No. 5352596
GENERAL INFORMATION:
APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.
TITLE OF INVENTION: Pseudotables Virus Deletion Mutants
TITLE OF INVENTION: Involving the EPO and LIT Genes
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSER: Curtis P. Ribando
STREET: 1815 NO. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P.
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4128
TELEFAX: 309-685-4011 ext.513
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDNESS: Double
TOPOLOGY: Linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
AMBI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudotables virus
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

alignment_scores:
Quality: 508.00 Length: 2060

Ratio: 0.647 Gaps: 104
Percent similarity: 38.107 Percent identity: 22.961

Alignment block:
US-09-508-821B-6 x US-07-945-283-1/rev ..

Align seg 1/1 to reverse of: US-07-945-283-1 from: 1 to: 8438

```

31 ArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAla 47
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7236 CGGCGCCATCTGGCGCGCATGCA.....AAGCAGCGGAGAGA 7193
47 aLYAspTYrTYrAsnProGlnProTYrProSerTYrGluGly..... 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7192 AGCGGGGAGAAAAATCCCATTCGCGCGCGCTCGCGGAAGTCCGCGCGA 7143
62 .....GlyAlaGlyThrProSerGlyThr 69
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7142 AAATCGGCGCATTCGTCGCTTACCTGGGGCGCGCTCTCTCGGGCGCT 7093
70 AlaAlaAlaValAlaAlaAspLYsTYrHisArgGly..... 81
||| |||||
7092 .....TATAAGCGCGTCTCCATCGTAGCAC 7067
82 .....SerLYsAlaLeuProThrGln..... 88
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7066 TTCACGCGGTGAGGTAGGACGACATCGTCTCTGCCAACCGAGGG 7017
88 ..... 88
7016 ATCCGATCGTCTCGCTCGCGCGGACTGTGAACACTCGGCTCTCGG 6967
89 .....GlnGlyLeuGlnLYsArgProAlaPheProGly.....Ty 100
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6966 CGGCTACGACCTCGACGAGCGCGCCGACCCAGAGGCTCTCGGCCAC 6917
100 rGlyValGlnAspSerSerProTYrProGlyArgTYr..... 112
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6916 CAGAGAGAGTCTCTCTCTCTCGCGCGCTCTCTCTCTCTCTCTCT 6867
113 .....AlaGly GluGlu..... 116
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6866 GCGCGCCGCTCTCGGCGCTCGCGCGCGCTCTCTCTCTCTCTCTCT 6817
116 ..... 116
6816 CCGATCGAGACCATCGCGACGATCTCTTTGACTTCATCGAGACCGAGG 6767
117 .....SerLeuGlnAlaTrpGlyAlaProGlnProProGlnPro 131
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6766 CAATTCAGCCAGCTCTG.....CGGCGCGCGCGCGCGCGCG 6726
132 .....GlnProLeuProAlaGlyValAlaLYsTYrAspGln 143
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6725 AGAGAGAGGATCGCGTCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 6684
143 uAsnLeuMetLYsTYrAlaValProProSerArg.....GlnTYrAlaG 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6683 .....GCTCCGCGCGCGCGCTCTCTCGCGGAGGATCTCTCTCGG 6641
159 luGlnLYsAlaGlnValProPheArgThrHisSerLeuHisValGlnGln 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6640 CCGGCGCGCTCTCTCTCG.....ACGAGC 6615
176 ProProProGlnGlnProLeuAlaTYrProLYsLeuGlnArgGlnLY 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6614 CCGCGAGCGCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 6579
192 sLeuGlnAsnAspIleAlaSerProLeuProPheProGlnGlyThrHisP 209
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6578 .....CCGAGGCCACGAGGCCGC..... 6561
209 heProGlnHisSerGlnSerPheProThrSerSerThrTYrSerSerSer 225

```

```

6560 ..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
226 ValGlnGlyGlyGlnGlyAlaHisSerTYr..... 236
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6584 GGCAGCGCGCGCGCTCGCGGAGATCGTGTGTAGACGACGAGAC 6484
237 LYsSerCysThrAlaProThrAlaGlnProHisAspArgPro..... 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6483 GAGGAGGAGGAGAGCGCGCTCTCTCGCGCGCGCGCTCTCTCTCGG 6434
251 .....LeuThrAlaSerSerSerLeuAlaPro 259
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6438 TCTCCATCAGCGCTCGCGACGCTACCTCGTCTCTCTCTCTCTCTCT 6384
260 GlyGlnArgValGlnAsnLeuHisAlaTYrGlnSerGlyArgLeuSerTY 276
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6383 GGGCGG.....GCTCGGGCGCGCGCGG..... 6360
276 rAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaL 293
6360 ..... 6360
293 euGlnSerArgHisAlaGlnGlnThrLeuHisTYrGlnAsnLeuAla 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6359 ..CGACCGCGCGACACTCGCAGCGG..... 6336
310 LYsTYrGlnHisTYrGlyGlnGln..GlyGlnGlyTYrCysGlnProAsp 325
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6335 .....CAGAGCGCGCGCGCGCGCGCGCGCGCGCTCTCTCTCTCT 6292
326 AlaAlaValArgThrProGlnGlnTYrTYrGlnPheSerProSerSe 342
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6291 GCAGCC.....TCCCGCGCGCTC 6275
342 rSerHisSerProAlaArgSerValGlyArgSerProSerTYrSerSer 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6274 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6225
359 hrPro.....SerProLeuMet.....ProAsnLeuGlnAsnPhe 370
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6224 CCGCGGAGATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6175
371 ProTYrSerGln.....GlnProLeuSerTh 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6174 CCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6125
379 rGlyAlaPheProAla..... 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6124 GCGCGAGCGCGCGACGCGCGCGCGCGCGCTCTCTCTCTCTCTCT 6075
384 ..... 384
6074 CCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6025
385 .....GlyLeuThrAspHisSerHisPheMetProLeu..... 395
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6024 GACCGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5975
396 .....LeuAsnProSerProThrAspAlaThrSerSer 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5974 GCGCGCGCGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5925
407 alaSPThrGlnAlaGlyAsnCysLYsProLeuGlnLYsAspLYsLeuPro 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5924 TCTCT..... 5920
424 GluAsnLeuSerAspLeuSerLeuGlnSerLeuThrAlaLeuThrIle 440
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5919 .....TCCACCCCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5882
440 uGlnValGluAsnIleSerAsnThrValGlnGlnLeuLeuSerLYsA 457
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

5881 CGCCGTGAAGCC..... 5869
457 IaaIaValProGlnIlySerGlyValIysAsnLeuValSerArgThrPro 473
5868GTCCGCCCGCCGCC 5854
474 GluGlnHisIlySerGlnHisCysSerProGluGlySerGlyTyrSerAl 490
5853 GAGAAAGACGAG.....GACGACTCGGCTCGCCGG 5822
490 agluProIaGlyThrProIeSerGlnProProSerSerThrProGlns 507
5821 GGAC...GGCGGGCGCCCTCGACAGAGACGCCGCCGCCGCGC..... 5782
507 erThrHisAlaGluProGlnIleuAlaAspTyrLeuSerGlySerGlnAsp 523
5761CGCGCGGAGAGAGC 5767
524 ProIeGluArgSerPheLeuTyrCysAsnGlnAlaArgIly..... 537
5766 GCCTTCGGCGCGCGCGCGCTTCAGACGCTCCAGACGCGCGCGAGCGA 5717
538SerProIaArgValAsnSerAsnSerIlyAla..... 548
5716 CTCGACCTCTCCCGCCGCTCCGCTCCGCCGCCGCCGCCGCCGCCGCG 5667
549LysProGluSerValSerThrCysSerVal 558
5666 CGCGCGCGCGCGCGCGCGCGCTCGCTCTCTCTCTCTCTCTCTCTCTCT 5617
559 ThrSerProAspAspMetSerThrIlySerAspAspSerPheGlnSerIle 575
5616 TCCT 5567
575 uHisGlySerLeuProIeAspSerPheSerIlySerValAlaGlyGln 592
5566 CG 5532
592 rGAspCysProArgLeuLeuLeuSerAlaLeuAlaGlnIleuAla 608
5531 CACGCTCTCGCGCG.....GGCGCGCGCGCGCGCGCGCGCGCTCGGCGC 5488
609 SerGluIleuLeuGlyLeuGlnIleuAlaIleGlyValIysAlaAspIlyAl 625
5487 TCCT 5450
625 atIPAlaGluAlaProSerLeuValIlyAspSerSerIlyProIlePhe 642
5449 CGCT 5400
642 erLeuGlnAsnHisSerAlaCysLeuAspSerValAlaIlySerAlaTrp 658
5399 CCACCAACAACAC.....CTCTCGCTCAGGCCGAGCGCGCGCGCGCG 5359
659 ProArgProGlyIleu.....ProGluI 666
5358 CGACCGACGAGCGCGCTCTCACCGCGCGCGCGCGCGCGCGCGCGCT 5309
666 atLeuProAspSerLeuGlnLeuAspIlyGlyValAsnAlaIlyAspPhe 683
5308 CGACCTTCGCGCGCGCGCGCGCTTCAGCGCGCGCGCGCGCGCGCGCTCC 5259
683 erProGlyLeuPheGlnAspProSerValAlaPheAlaThrProAspPr 699
5258 GCGAGGCGCTCTGGGAGAGAGAGAGCTCGCGCGCGCGCGCGCGCGCTAC 5209
699 OlyIlyThrThrGlyProLeu.....SerPheGlyThr 711
5208 CGCGCGCGCGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5159
711 ys.....ProThrLeuGlyValPro 717
5158 GAGGAGACGACGAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5109

718 Ala...ProAspProThrThrAlaAlaPheAspCysPheProAspThr 733
5108 GCGAGCGCATGTCTGCTGAGAGACCGCGGAGAGCGCGCGCGCGCGAG 5059
733 rAlaIaSerSerAlaAspSerAlaAspProPheAlaTrpProGlu 749
5058 CGCTTCACACATCTCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 5012
750 AsnLeu.....GlyAspAlaCysProArgTrpAlaIleuHisProG 763
5011 CGCGCT 4962
763 yGluLeuThrIlyGlyLeuGlnIleuGlyIlyAlaSerAspGlyTyr 780
4961 AGCGCATGGCGCGCGAGACCGCTCTGCGCGCGCGCGCGCGCGCGAG 4912
780 erIlyGly.....AspThrHisGln 786
4911 GCGGTGGCATGAGCGCGCGCTACAGATCCGACGAGAGACCTTCACTCT 4862
787 AlaSerAlaCysLeuGlyPheGlnIleuGlnAspProProGlyGlyVal 803
4861 GCA.....GAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4836
803 lAlaSerLeuProGlyAspPheIysGlnGlnIleuAlaGlyIlyAla 819
4835 ACATGGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 4786
820 GluGlnAlaGlyIleuLeuGlnCysProGlnVal..... 831
4785 GTGAGAGCG.....CTCTGCGCGCGCGCGCGCGCGCGCGCGCTCT 4748
832 ...AlaIlyAlaAspArgTrpLeuGlnAspSerArgHisCysCysSer 847
4747 GCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4658
847 hrAlaAspPheGlyAspLeuProLeuLeuProThrSerArgIlyGln 863
4697 TGGCTGCGCGCGCGCGCGCTCTGAGCGCGCGCTCTCTCTCTCTCT 4663
864 AspLeuGlnAlaGlnIleuGlyIlySerSerLeuCysGlnLeu...Leu 879
4662TCCCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4646
879 ySerProGlnIleuArgProGlyMetGlnAspProLeuSerProIlyAla 896
4645 CG 4612
896 rGluLeuIleuGlyThrIlyGlnIleuAlaGlnIleuValLeu.....Asp 909
4611GCTGCTCTGAGAGGTGACCGCGCGCTCTCTCTCTCTCTCTCT 4567
910 SerIlyAlaGlyTyrPhe...SerProCysHisLeuSerGlyIleuSer 925
4565 GCGATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4517
925 lIleuLeuGlyProThrVal.....G 933
4516 CGCT 4467
933 lYThrGlnSerIlyValGlnSerTrpPheGlnIleuSerLeu...SerHis 949
4466 GCT 4417
949 lYsProGlyGlnIleuGlyProAspGlyIleuArgAlaPro..... 962
4416 CGCGCGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4367
963GlyAspSerThr 967
4366 GCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4317

1223 LysArgLys..... 1225
1226 1228
1229 1231
1230 1232
1231 1233
1232 1235
1233 1237
1234 1239
1235 1241
1236 1243
1237 1246
1238 1251
1239 1254
1240 1259
1241 1264
1242 1269
1243 1274
1244 1279
1245 1284
1246 1289
1247 1294
1248 1299
1249 1304
1250 1309
1251 1314
1252 1319
1253 1324
1254 1329
1255 1334
1256 1339
1257 1344
1258 1349
1259 1354
1260 1359
1261 1364
1262 1369
1263 1374
1264 1379
1265 1384
1266 1389
1267 1394
1268 1399
1269 1404
1270 1409
1271 1414
1272 1419
1273 1424
1274 1429
1275 1434
1276 1439
1277 1444
1278 1449
1279 1454
1280 1459
1281 1464
1282 1469
1283 1474
1284 1479
1285 1484
1286 1489
1287 1494
1288 1499
1289 1504
1290 1509
1291 1514
1292 1519
1293 1524
1294 1529
1295 1534
1296 1539
1297 1544
1298 1549
1299 1554
1300 1559
1301 1564
1302 1569
1303 1574
1304 1579
1305 1584
1306 1589
1307 1594
1308 1599
1309 1604
1310 1609
1311 1614
1312 1619
1313 1624
1314 1629
1315 1634
1316 1639
1317 1644
1318 1649
1319 1654
1320 1659
1321 1664
1322 1669
1323 1674
1324 1679
1325 1684
1326 1689
1327 1694
1328 1699
1329 1704
1330 1709
1331 1714
1332 1719
1333 1724
1334 1729
1335 1734
1336 1739
1337 1744
1338 1749
1339 1754
1340 1759
1341 1764
1342 1769
1343 1774
1344 1779
1345 1784
1346 1789
1347 1794
1348 1799
1349 1804
1350 1809
1351 1814
1352 1819
1353 1824
1354 1829
1355 1834
1356 1839
1357 1844
1358 1849
1359 1854
1360 1859
1361 1864
1362 1869
1363 1874
1364 1879
1365 1884
1366 1889
1367 1894
1368 1899
1369 1904
1370 1909
1371 1914
1372 1919
1373 1924
1374 1929
1375 1934
1376 1939
1377 1944
1378 1949
1379 1954
1380 1959
1381 1964
1382 1969
1383 1974
1384 1979
1385 1984
1386 1989
1387 1994
1388 1999
1389 2004
1390 2009
1391 2014
1392 2019
1393 2024
1394 2029
1395 2034
1396 2039
1397 2044
1398 2049
1399 2054
1400 2059
1401 2064
1402 2069
1403 2074
1404 2079
1405 2084
1406 2089
1407 2094
1408 2099
1409 2104
1410 2109
1411 2114
1412 2119
1413 2124
1414 2129
1415 2134
1416 2139
1417 2144
1418 2149
1419 2154
1420 2159
1421 2164
1422 2169
1423 2174
1424 2179
1425 2184
1426 2189
1427 2194
1428 2199
1429 2204
1430 2209
1431 2214
1432 2219
1433 2224
1434 2229
1435 2234
1436 2239
1437 2244
1438 2249
1439 2254
1440 2259
1441 2264
1442 2269
1443 2274
1444 2279
1445 2284
1446 2289
1447 2294
1448 2299
1449 2304
1450 2309
1451 2314
1452 2319
1453 2324
1454 2329
1455 2334
1456 2339
1457 2344
1458 2349
1459 2354
1460 2359
1461 2364
1462 2369
1463 2374
1464 2379
1465 2384
1466 2389
1467 2394
1468 2399
1469 2404
1470 2409
1471 2414
1472 2419
1473 2424
1474 2429
1475 2434
1476 2439
1477 2444
1478 2449
1479 2454
1480 2459
1481 2464
1482 2469
1483 2474
1484 2479
1485 2484
1486 2489
1487 2494
1488 2499
1489 2504
1490 2509
1491 2514
1492 2519
1493 2524
1494 2529
1495 2534
1496 2539
1497 2544
1498 2549
1499 2554
1500 2559
1501 2564
1502 2569
1503 2574
1504 2579
1505 2584
1506 2589
1507 2594
1508 2599
1509 2604
1510 2609
1511 2614
1512 2619
1513 2624
1514 2629
1515 2634
1516 2639
1517 2644
1518 2649
1519 2654
1520 2659
1521 2664
1522 2669
1523 2674
1524 2679
1525 2684
1526 2689
1527 2694
1528 2699
1529 2704
1530 2709
1531 2714
1532 2719
1533 2724
1534 2729
1535 2734
1536 2739
1537 2744
1538 2749
1539 2754
1540 2759
1541

7759 CTCACACAGCCGCGCTCTGTCCGGATGGCCCCCCTTCGCTGCCTTGGCGGCG 7808

116 luSerLeuGlnAlaTrpGlyAlaProGlnProProGlnPro... 131
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

7809 GCTCGCGCTGCTCTCGAGCACGACGAGACACACGTCACGTCGCCGACA 7858

132GlnProLeu...ProAlaGln 137
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

7859 TCGACTGGGGCGCTTTGGCGCTTCGTCAGCGCGCGCTGCCCGCGCC... 7906

137 yValAlaIstYrAspGluAenLeuMetIstYsThrAlaValProProS 154
yValAlaIstYrAspGluAenLeuMetIstYsThrAlaValProProS 154
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

7907CGTCTCGCGCATTGCCCGCA 7928

154 exArgGlnTyAlaGlnGlnGlyAlaGlnValProPheArgThrHisser 170
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

7929 GCGCAGCGCGCTCTCGAGACCACGAGGCG... 7960

171 LeuHisValGlnGlnProProproGlnGlnProLeuAlaTyProLy 187
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

7961CGTCTCGCGCATTGCCCGCCC... 7984

187 sLeughnArgLysLeuGlnAenAspIleAlaserProLeuProPheP 204

7984 7984

204 roGlnGlyThrHisPheProGlnHisserGlnSerPheProThrSerSer 220
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

7985CGCAGCTCTCTCG 7996

221 ThrTyrserservalGlnGlyGlyGlnGlyAlaHisserTyLy 237
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

7997 ACAAGCTCGGAGCGCGCTCGGAGACGAGCAGCTCGTCTGCTCTCTCG 8046

237 sSerCys.....ThrAlaProThrAlaGlnProH 247
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8047 CTGTCGCCACCGAGACGCCGCTGCTCGGCCACGAGCGCCTCCCA 8096

247 IsAspArgProLeuThrAlaserSerSerLeuAlaProGlyGlnArgVal 263
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8097 TGTCGACCCGACAGGGCTTCTCTGATC... 8125

264 GlnAsnLeuHisAlaTyGlnserGlyArgLeuSerTyRspGlnGln 280

8125 8125

280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeuGlnSerArgH 297

8125 8125

297 ishIsalagInlThrLeuHistyrGlnAsnLeuAlaLystyrGlnHis 313
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8126TCGGTCTCGATTGGCTCA 8143

314 TyrglyGlnGlnGlyGlyTyrcysGln.....ProaspAlalaVa 328
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8144 TGCGCGTAGCTTCGCGCGCTGTGACAGCGCCACCGCATCAAGCTC 8193

328 lArgThrProGlnGln.....TyrrGlnThrPheSerP 340
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8194 CCGGCCACCTCGCTTCGACCAICCTCTCTCTCATCGAGTCCGCTT 8243

340 rOserSerHisSerProAlaArgSerValGlyArgSerProSerTy 356
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8244 TTTCGCGACTCGCTCGCCAGCGCTTCGACAGAGGCTCTTCGCTG 8290

357 SerSerThrProProLeuMetProAsnLeuGlnAenPhe...ProTy 372
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8291AGCCGAGCGCGCGGCTCCCGGCGCTTCGCGCGCG 8328

372 sSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyLthrAspH 389
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||

8329 AGCAGAGCGCCATCGCATCTCGGATCGGATCGGCTTCGCTTCGCGCGCG 8378

[illegible]

10769 TCAGCCCTCGGAGAGCTCGAGCGCTACCGCTGATCCCGCTGCTC 10818
1152 hrGlnGlnIlePheHisSerLysArgArgArgProSerGluGlyArgLeu 1168
10819 GG.....CTCCATTGAGAGAGAGGACCTCG..... 10849
1169 ProAnCys.....ArgAl 1173
10850 CCGCGCTGCTCTCTCTCTGGGGGAGCTCTCTACCGAGGCGCTCGCTC 10899
1173 arThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190
10900 GACTGGAGGAGCTTCTTCGCGCGCTACGCTCCCGGCAAGGTCCCTCC 10949
1190 erSerSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203
10950 CACCTACCCCTTCAGGAGAGCGGTTCCTGGCTCGAGCTCTCCAGGAGC 10999
1204 ArgVal..... 1205
11000 AAGCGTTCGAGCTGCGCTCGGAGGCTGACCTCGGCGGACCAATCCG 11049
1206ProLysProGlyAla..... 1210
11050 CTGCTCGGCGCGCGCTCGCTTCGCGACCGCGGTGGCTTCTTTTAC 11099
1211GlySerLysLeuSerAspArg 1217
11100 AGGCGGCTCTCTCTCGAGAGCACCCTGCTGCTCGAGGCGCATCGCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGCACACCATCTACCGGCGACCGCTTCCTCGAGCTCGCCCTGCAC 11199
1229MetAlaProValProThrLys.....LysArgA 1238
11200 GTGCGCCACCGCTCGCGCTTCGACACCTCGAGAGCTCACGCTCGAGGC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254
11250 CCCTCTCGCTCTCCCATCAGCAGACCGTCTCTCTCGAGATCTCGCTG 11299
1255 Gly..... 1255
11300 GGCCCTGGAGCAGCAGCAGAGAGGCGCTCTCTTTCCATAGCCGACAA 11349
1256GluMetG 1258
11350 GAGGAGCGCTTCAGGAGCGCCCTGGACTCGCCACGCCAGCGGCTCTCT 11399
1258 lyArgArgArgGlyLeuArgValProProProSerArgGlyCysLeu 1274
11400 CTCGCGCGGCGCCCATCTCTCGCGCGATCTCCAGAGTGC..... 11443
1275 LeuProArg.....LysProSe 1280
11444 CTCCTCTGAGTCCATCCGCTGGACCTCGAAGGCTCTACGCAACCCCTC 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCGACCTCGGCTGCTACGCGCGCGAGTCCAGGCGCTCGCTCGCT 11543
1290 InSerSer.....His 1293
11544 CTACAGCGCGGAGAGCTCTTTGCGAGACCAAGCTCCCGGAGCGG 11593
1294 ProArgArg.....ProProPheLeuProGlnAlaArgLeuSerAlaAlaP 1309
11594 CCGAAAGGATCGCCCGGTTCGCC...TCCACCTCGGCTGCTCGAC 11640
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgGlyArgGlyL 1326
11641 AGCCCTCTGATGCTGCTGGCTTTGAGGAGCAGGAGGAGGCTGCTC 11690

1326 eu..... 1326
11651 TCTGCCCTTCTGTCGAGCGAGTCTCGTGGCTCGCTCGGTGGCCACCA 11740
1327LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339
11741 CTTTGGCGCTGGCTTCACCGCTCCCAAGGTGAATCTCGCTCTGATC 11790
1340LeuLysLysPheA 1344
11791 GTCCTGGCGAGCGCGAGGTGACCTCTTGCCTCGGTGCAGCGCTGC 11840
1344 lacys..... 1345
11841 GATCGGAGCAGCTCCGCGCGAGCTCGGACCCCGGAGCTTCOCACC 11890
1346LysAlaPro..... 1348
11891 ATGATGCGCTTCTTCGCTCGTCACTGGAGGAGCTCCAAAGCCCACTTCA 11940
1348 1348
11941 CCGCTCGCGCGCGAGCGGTCTTCTCGGCACAGCGGCGCACGATCT 11990
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
11991 CGCGCTCGAGCGCGCTCGCGCTACCGGAGCTCGCTCGCTCCGAA 12040
1361 SerAspLysArgGlyLeuLys...GlyAlaGlyGlySerProValG 1376
12044 GCGCGCTCGACAGGCGCTTCGCTCCGCGCTCGCTCGCTCGCC..... 12085
1376 yValGlnGluGlyLeuValAsnValGlyThrGlnLysLeuProThr. 1392
12086CCTTCATGATCGACCGCGAGCGACCTCTCGCGAGCGCCACGA 12131
1393SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404
12132 GCGCACCGCTCGACCTCGCTTCGCAAGCTGGCTCGCGGAGAAC 12181
1405 ArgSerLeuLysLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421
12182 GCCTCGCTCGCTCGCGCTCTCTCGTCCCGGAGCGCGCTCGCCACC 12231
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438
12232 CACACCGAGAGCGCTCAAGACTCG.....CTCAGCGCGCTCTG 12275
1438 lyThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454
12276 GGGCTCGCGCTCGCGCAAGTGAGACCCAGACCTCCGCTCTTCC 12325
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrLeuGlyProAl 1471
12326 TGTGTCAGATCG.....ACCTCAGGAGCGCTCCACGAGCGCTG 12366
1471 alenLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 CTAGGCGCTCGACAGAGAGCGGCTCGCTCCATTTCCAGCAAGGACCTTC 12416
1488 luAspAsnSerGlyGly..... 1493
12417 ACCCTCATCTCCGAGGTTGGGCAACCAAGCTCGAGCGCGCTCATCC 12466
1494GlyGlyLys.....LysProLysMetGluG 1502
12467 CGCGCGAAGCACCGAGTGGCGCTCCATTTCCAGCAAGGACCTTC 12516
1502 uLeu...GlyProAlaSerGlnProProGlnGlyArgProCysGlnProG 1518
12517 GACGCGCTCGCTCTCGTACGCGCGG.....AGCGCGAGGCGCGCT 12560


```
seq_name:/cgn2_6/prodata/2/ina/5a_COMB.seq:US-08-456-837-6
seq_documentation_block:
1 Sequence 6, Application US/08456837
2 Patent No. 5643774
3
4 GENERAL INFORMATION:
5
6 APPLICANT: Schupp, Thomas
7
8 APPLICANT: Ligon, James M.
9
10 APPLICANT: Beck, James Joseph
11
12 APPLICANT: Hill, Dwight Steven
13
14 APPLICANT: Ryals, John Andrew
15
16 APPLICANT: Gathney, Thomas Deane
17
18 APPLICANT: Lam, Stephen Ying
19
20 APPLICANT: Hammer, Phillip E.
21
22 TITLE OF INVENTION: Genes for the synthesis of
23
24 TITLE OF INVENTION: antipathogenic substances
25
26 NUMBER OF SPOUNDS: 22
27
28 CORRESPONDENCE ADDRESS:
29
30 ADDRESSEE: Ciba-Geigy Corporation
31
32 STREET: 7 Skyline Drive
```

```

; MOLECULE TYPE: DNA (genomic)
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
;
US-08-456-837-6

```

US-09-508-821B-6 x US-08-456-837-6

Align seg 1/1 to: US-08-456-837-6 from: 1 to: 28958

7609 GCGGGGACACGCAAGCGGCTACGGGCGCGAAGCCCTTCCTCGAGC 7610

7559 CCTGCCGAGCATCGCGCGAGCGCTGATTGACAGCGACGTCGGTGGCT 7708

7709 GGGCGCGTGGGGCGGGCGGCGCATGGCCACCGATCAGCGGCGAGCCAC 7758

7759 CTCACACGCGGCTCTGTGCGGAGTGGCCCCCTCGCTTGCCTGGCGC 7808

7809. GCTGGCGCTGGCTCTGGAGCAGCAGAGACCACCGTCAACCGTCCGACA 7858

7859 TCGACATGGGGCGCGCTTTGCGCCATCGTTCAGCGCGCGCTCGCCCCCGCC.. 7906

7907CGCTCTCGCGGATTTCGCCGA 7928
154 eArgGlnTyrAlaGlnGlnGlnAlaGlnValProPheArgThrHisSer 170
7929 GCGCAGAGCGCTCTCGAGACGAGCGG..... 7960
171 LeuHisValGlnGlnProProProGlnGlnProLeuAlaTyrProLy 187
7961CGTCTCGAGCATGCGCGGCC..... 7984
187 sLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerProLeuProPheP 204
7984 7984
204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProHisSerSer 220
7985CCGACCTCTCTCG 7996
221 ThrTyrSerSerValGlnGlnGlyGlyGlnGlnAlaHisSerTyrLy 237
7997 ACAAGCTCGGAGCGCTCGGAGAGCGAGCATTCGTCTGCTCTCG 8046
237 sSerCys.....ThrAlaProThrAlaGlnProH 247
8047 CTGGTGGCCACAGAGCGCTCTCTCGGCGACGAGCGCTCTCGCA 8096
247 isAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgVal 263
8097 TGTGACCCGACAGAGGCTCTCTCGATC..... 8125
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280
8125 8125
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeuGlnSerArgH 297
8125 8125
297 isHisAlaGlnGlnThrLeuHisTyrGlnLeuAlaLysTyrGlnHis 313
8126TCGGTCTCGATTGCGTCA 8143
314 TyrGlyGlnGlnGlnGlnGlyTyrCysGln.....ProAspAlaAlaVa 328
8144 TGGCGCTGAGCTTCGCGGCGCTGCAACAGGCGACCGCATCAAGTC 8193
328 lArgThrProGlnGln.....TyrTyrGlnThrPheSerP 340
8194 CCGGCGACCTCGCTTCGACCATCCCTCTCTCATCGAGTCGCGCTCTT 8243
340 roSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
8244 CTGGCGGACTCGCTCGCCCGCGCTCGGCGGCTTCGCGCGCG 8290
357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...ProTy 372
8291AGCCGCGCGCGCGCTCGGCGGCTTCGCGCGCG 8328
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389
8329 AGGAGGAGCCATCGCATCGTCGTCGATGCGCTCGGCTCGCGCGG 8378
389 isSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404
8379 CGTGGCGGATGTCAGCTCTTTGGAGATTCTTGGCCCGAGGAGCGGAGC 8428
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420
8429 GCGTCG.....AGCCATTCCAAAGCCCGA 8454
420 pLysLeuProGlnAsnLeuSerAspLeuSerLeuGlnSerLeuThrA 437
8455 TGGGATGCGGCTGCTCTAGACCCGACCGCCGACCAAGCAAGAG 8504

437 laLeuThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeuLeu 453
8506 CTAGCTCGGCGATGCCGCATCTCGACAGGTCTGACCTCTCGACCTG 8554
454 LeuSerTyrAlaAlaValProGlnLysLysGlyValLysAsnLeuValSe 470
8555 CTTCTTTGGCATGACCCCGGAGGAGCCAAAC.....AC 8589
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGluGlySerG 487
8590 CTCGACCCGC.....AGCACCGCTCTCTCTCTCTCTCTCTCTCTCT 8618
484 lyTyrSerAlaGluProAlaGlyThrProLeuSerGluProPro..... 501
8619 ATCTGCTGGCAGGCTCTCGAAGACCGCGCATCTCTCTCTCTCTCTCT 8668
502SerSerThrProGlnSerThrHi 509
8669 AGGATTCCCCCACCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8718
509 alaGlu.....ProGlnGluAlaA 516
8719 TTGGAGAGGCGAGCAGCAGAGATTCCGACGCTTATGCTCTCAAGGCA 8767
516 spTyrLeuSerGlySerGluAspProLeuGlnArgSerPheLeuTyrCys 532
8768CCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8803
532 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLy 549
8804 CGCTCGGCTCTCAAGGCGCG.....CGCTCTCTCTCTCTCTCTCTCTCTCT 8850
549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSert 566
8851 TCCTCGCT 8900
566 hrLysSerSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
8901 GTGCAACCT 8950
583 SerPheSerLysPheValAla..... 589
8951 GCTTCT 9000
590GlyGluArgAspCysP 595
9001 AAGACCT 9050
595 roArgLeuLeuLeuSerAlaLeu.....AlaGlnGluAspLeuAla 608
9051 GGTCT 9100
609 SerGluLeuLeuGlyLeuGlnGlnAlaIleGlyGluLysAlaAspLysAl 625
9101 TCCT 9150
625 a.....TrpAlaGluAlaProS 631
9151 GGTATCACCGCCCGCAAGCGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9200
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643
9201 GGTCT 9250
644GluAsnHisSerAlaCysLeuAspSerVa 653
9251 ATGGCACCGGACCT 9289
653 lalaLysSerAlaTyrProArgProGlyGluProGlnAlaLeuProAspS 670
9290AGCCCT 9325

670 erLeuGlnLeuAspLysGlyValAlaLysAspPheSerProGlyLeu 686
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThr 703
6930 CGCTCAAGACCAACATCGGCATCGAGGCGCCCTCGGCGCGGCG 9399
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValAlaPro 720
720 SPProThrThrAlaAlaPheAspCysPhePro.....AspThr 732
733 ThrAlaAlaSerSerAlaAspSerAlaAspProPheAlaThrPro 749
749 uAsnLeuGlyAspAlaCysProArgThrProGlyLeuHisProGly 766
766 hrLysGlyLeuGlnGlnGlyLysAlaSerAspGlyLysSerGly 782
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGln 795
795 uGluAspProProGlyGlyLysValAlaSerLeuProGlyAspPhe 812
812 InGlnValAlaGlyValLysGlnGlnGlnGlnGlnGlnGlnGln 828
829 ProGlu.....ValAlaLysAlaAspArgThrPheGln 840
840 PserArgHisCysCysSerThr.....AlaAspPhe GlyAspLeuPro 854
855 LeuLeuProProThrSerArgLysGlnAspLeuGlnGlnGlnGln 871
871 rSerSerLeuCysGlnLeuLeuGlySerProGlnGlnArgProGly 888
888 LAspProLeuSerProLysAlaProLeuLeuLeuLeuLeuLeu 904
905 GluGlnValLeuAspSer.....LysAla.....GlyThr 914
914 PglYserProCysHisLeuSerGlyGlnSerValIleLeuLeuGly 921
921 hrValGlyThrGlnSerLysValGlnSerThrPheGlnSerLeuSer 947
948 HisMetLysProGlyGlnGlnGlyProAsp.....GlyGlnArgAla 962

10651AGTGAAGCCTGCTGCGCGCGCGCGCGCGCGCGCG 10101
962 oGlyAspSerThrSerAspAla..... 970
10102 CCTCCCTGACCGGCTGCGAGCTGACAGCCCGCTCTTTCGCGAT 10151
971SerLeuAlaGlnLysProAsnLysProAlaVal 981
10152 GGTCTCCGTCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCG 10201
982 ProGlnAlaProLeuAlaLysGlnGlnProValProArgLysSer 998
10202 TCAGCGGACAGAGCGAGGAGAGATGCGCGCGCTGCTGCGAGGCG 10251
998 uArgSerArgArgValHisArgGlyLeuProGlnGlnGlnAspSer 1015
10252 CTCCTCCCTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGln 1028
10287 GCGAGAGAGAGCGCTGACACCGTCCGCGCGCGCGCGCGCGCG 10336
1029 SerCysThrGlyProProGlnGln.....MetGlnGlyAla.. 1041
10337 TCAGCGGAGCGCTGCGACCTGACAGCTGACCTGCGCGCGCGCG 10386
1042GlyAlaProGlyArgGlyAlaSerGln 1051
10387 AGGCTTCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10436
1051 LysLeuProArgMetLysThrArgSerLeuThrAlaLeuSerGln 1067
10437 CGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10486
1068 ...ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080
10487 TGTTCGCGCGCAAGATCCGCGTGGACTGACGCTGCGCGCGCG 10536
1081ProAspLysLeuGly 1086
10537 GAGCGCGTCCAGAGAGCTGCGCGAGGTCTGACAGCAATCGCT 10586
1086 LysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLysPro 1102
10587 GAGCTGCGAGCTGCTCTTATT.....CAACGCTGA 10618
1103 ProLysAlaAlaSer...SerProSerAsnProAla..... 1113
10619 CGGCGACAGAGCTGACAGGCTCGAGCTGAGCGGCGTACTGATGA 10668
1114AlaLeuProValAlaSerAspSerProMetGlySerLys 1128
10669 AACCTCCGCAACCGTCTGTTCTGAGCGAGCGAGCGGCTCTGA 10718
1128 hrLysGlnThrAspSerProSer..... 1135
10719 CGATGGCAATCGCTTCTGAGAGTCAAGCCCATCCGCTGATCGC 10768
1136 ThrProGlyLysAspGlnArgSerMetLeuArgSerArgThrLys 1152
10769 TCAGCGCGCGCGAGAGCTGAGGCGCTCAACGCTGATCCGCTG 10818
1152 hrGlnGlnLysPheHisSerLysArgArgArgProSerGlnGln 1168
10819 GG.....CTCCATTGACAGAGAGAGCGCTGCTG..... 10849
1169 ProAsnCys.....ArgAla 1173
10850 CCGCGCTGCTCTGCTGCGCGAGCTGCTGACCGAGGCGCGCGCT 10899
1173 aThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLys 1190

10900 GACTGGAGGAGCTCTTCGGCGCCTACGCTCCCGCAAGGTCTCCCTCC 10949
1190 erSerSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203
10950 CACCTTACCCCTTCAGGAGAGCGGTCTTGGCTCGACGCTCTCCACGAGC 10999
1204 ArgVal..... 1205
11000 AACGCTTCGACGCTCGCTTCGCGCCTGACCTCGCGCGCAATCCG 11049
1206ProLysProGlyAla..... 1210
11050 CTGCTCGCGCGCGCTCGCTTCGCGAGCGCGGTCTTCTCTTTAC 11099
1211GlySerLysSerAspArg 1217
11100 AGGGGGCTTCCTCGCGAGACCGCTGGCTCGAAGGCCATCCGCTCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGCACACCCATCTTACCGGCACCGCTTCGAGTTCGAGCTCGCCCTCAC 11199
1229MetAlaProValProThrLys.....LysArgA 1238
11200 GTGCGCCACCGCTCGCGCTCGACACCGCTCGAAGAGCTCACGCTCGAGGC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerSerSerSerSerSer 1254
11250 CCTCTCGCTCTCCATCGCAGACACCGCTCTCTCCAGATCTCCGCTCG 11299
1255 Gly..... 1255
11300 GGCGCTGGAGCGAGCAGACGAGGCGCTCTCTTTCCATAGCGCA 11349
1256GluMetG 1258
11350 GAGGACGCGCTTCAGGATGGCGCTCGACTCGCGCAGCGGCTCTCT 11399
1258 LysArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274
11400 CTCGCGCGGACCGCCATCCCTCTCGCGGATCTCCACAGTGGC..... 11443
1275 LeuProArg.....LysProSe 1280
11444 CTCCTCTGAGTGCCATCCCGGTGGACCTCGAAGGCTCTACGCAACCTC 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCCAACCTTCGCGCTTGCCTACGCGCGGAGTTCCAGGCGCTCCGCTCGT 11543
1290 InSerSer.....His 1293
11544 CTCACGCGCGGCGAGAGCTCTTTCGCGAAGCGCAAGCTCCCGGAAGCG 11593
1294 ProArgArg...ProProPheLeuProGlnAlaArgLeuSerAlaLaph 1309
11594 CGAAGAGATGCGCGCGGTTCGCC...TCACCTCGCGCTCGTGAC 11640
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgLyl 1326
11641 AGCGCGCTTCGATGCATCGCTTTCGAGGACGAGAGAGGAGCGTCCG 11690
1326 eu..... 1326
11691 TGTGCTTCGTGGAGCGAGTCTCGCTGCGCTCGGTGCCACCA 11740
1327LysLeuGlnAlaIleValGlnLysIleThrSerProser..... 1339
11741 CCTTGGCGGTGCGCTTCCACGCTCCAGAGGTGATCTCTCGTCTCGATC 11790
1340LeuLysLysPheA 1344
11791 GTCTGTGCGGACGCGCAGGTGACCTCTTTCGCTCGGTGCAAGCGCTCGC 11840

1344 laCys..... 1345
11841 CATGCGGAGAGCTCCGCGCGAGCTCCGACACCGCGAGCTTCCACC 11890
1346LysAlaPro..... 1348
11891 ATGATGGCTCTTCGCGCTGACTGGAGGAGCTCCAAAGCCCACTTCA 11940
1348 1348
11941 CGGCTCGCGCGCGCGGCTCTCTTCGGCACAGGCGGCCACGATCT 11990
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
11991 GCGGCTCGAGCGCGCGCTCGCGCTAGCGGAGCTCGCTGCCCTCCGAA 12040
1364 SerAspLysAspArgGlyLeuLys...GlyAlaGlyLysSerProValG1 1376
12041 GCGCGCTCGACGAGGCGCTCGCTTCGCGCGCTCTCGCTCGCC... 12085
1376 yValGlnGlnGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392
12086CCTTCATGATCGACCGCGGAGCGGAGCTCTCCGCGGCCACGA 12131
1393SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404
12132 GGCACCGCGCTCGCACTCGCTCTTGAAGCGCTGCTCGCGGAGAAC 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421
12182 GCTTCGCTCGCGCGCTCTCTCTCGTACCGCGCGCGCTCGCCACC 12231
1421 pCysPheLysThrGlnAlaPheThrSerProGlnAlaLeuGlnProGlyG 1438
12232 CACACCGAAGACGCTCAAGGAGCTCG.....CTCAGCGCGCTCTG 12275
1438 LysThrAlaLeuAlaProLysArgSerArgLysGlyArgAlaGlyAla 1454
12276 GGGGCTCGCGCTCGCGCAAGTGAGCAGCCAGACCTCCGCTCTTCC 12325
1455 HisGlyLeuSerLysGlyProLeuGlnLysArgProLysLeuGlyProAl 1471
12326 TGTTCAGATCG.....ACCTCAGGAGGCTCCCGAGGCGCTG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 CTAGCGCGCTCGACAGAGGAGCGGAGCTCGCTCGCGCAAGGAA 12416
1488 LysAsnSerGlyGly..... 1493
12417 ACCCTCATCGAGGTGGCGAACCGACCGCTCGAGCGCGCTCATCC 12466
1494GlyGlyLys.....LysProLysMetGluG1 1502
12467 CGCGGAGACCGACCGCTGGCGCTTCATATTCGACCAAGGCGCTTC 12516
1502 uLeu...GlyProAlaSerGlnProGlnGlyArgProCysGlnProG 1518
12517 GAGCGCTCGCGCTCGCGCGCGG.....AGCGCGCGCGCT 12560
1518 LysThrAlaGlnLysGlnProGlyHisThrAsnThrSerSerThrSer 1534
12561 CGCACAGCGCGCAAGTCCGCTCGCGTCCGCGCGGAGGCTCAACTTC 12610
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
12611 GCGATGCTCGACACCTTGGCATGTATCGGGCGGAGCGCGCGCG... 12658
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561
12659TCGAGCGGAGCGCGCGGCGATCGTTACTGAGT 12692

[illegible]

7984 7984
204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSerSer 220
|||||
7985 CGAAGCTCTCTCG 7996
221 ThrTyrSerSerSerValGlnGlnGlyGlyGlnGlnAlaHisSerTyrLy 237
|||||
7997 ACAAGCTCCGAGCGCTCGAGAGGAGGAGGAGCTCTCTCTCTCTCTCTCT 8046
237 sSerCys ThrAlaProThrAlaGlnProH 247
: |||
8047 CTGTGGCGCCAGACAGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8096
247 isAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnArgVal 263
: |||
8097 TGTCGACCCGACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8125
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnG 280
8125 8125
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 297
8125 8125
297 isHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 313
: |||
8126 TCAGTCTCTGATTCGCTCA 8143
314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328
8144 TGGCGCTCGAGCTGCGCGGCTGTCGACAGGCGCCGCGCAGCTCAAGCTC 8193
328 lArgThrProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 340
: |||
8194 CGGGCCACCTCGCTCTGACACATCCCTCTCTCTCTCTCTCTCTCTCTCT 8243
340 rOserSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
: |||
8244 CTTCGGGAGCTCTGCGCCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 8290
357 SerSerThrProSerProLeuMetProAsnLeuGlnGlnGlnGlnGln 372
8291 AGCCCGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8328
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389
: |||
8329 AGCGAGAGCCATCGCATCTGCGCATGCGCTCTCTCTCTCTCTCTCTCT 8378
389 isSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404
: |||
8379 CGTCGGCGATGCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8428
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnGlnGln 420
8429 GGGTCG AGCCCGATTCGACAGGCGCGG 8454
420 pLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThr 437
8455 TGGGATCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8504
437 lAlaLeuThrLeuGlnValGlnAsnIleSerAsnThrValGlnGlnLeu 453
8505 CTACGTCTCGGATCGCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8554
454 LeuSerTyrAlaAlaValProGlnGlnGlnGlnGlnGlnGlnGlnGln 470
8555 CTTCTTGGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8589
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGlnGlySerG 487
8590 CTGACCGCC AGCAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 8618

487. lyTyrSerAlaGluProAlaGlyThrProLeuSerGluProPro 501
ATTCCT 8618
8619 ATTCCT 8668
502 SerSerThrProGlnSerThrHi 509
8669 AGATTCCT 8718
509 salaGlu ProGlnGluAla 516
8719 TGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8767
516 spTyrLeuSerGlySerGluAspProLeuGlnGlnGlnGlnGlnGln 532
8768 CGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 8803
532 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAla 549
8804 CGCT 8850
549 sProGlnSerValSerThrCysSerValThrSerProAspAspMetSer 566
8851 TCT 8900
566 hrlYSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
8901 GTCT 8950
583 SerPheSerLysPheValAla GlyGluArgAspCysP 595
8951 GCT 9000
590 TrrAlaGluAlaProS 631
9001 AGACCT 9050
595 roArgLeuLeuLeuSerAlaLeu AlaGlnGluAspLeuAla 608
9051 GTCCT 9100
609 SerGluLeuLeuGlyLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 625
9101 TCCT 9150
625 a TrrAlaGluAlaProS 631
9151 GGTATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9200
631 erLeuValLysAspSerSerLysProProPheSerLeu 643
9201 GCT 9250
644 GluAsnHisSerAlaCysLeuAspSerVa 653
9251 ATGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9289
653 lAlaLysSerAlaThrProArgProGlyLeuProGluAlaLeuProAsp 670
9290 AGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9325
670 erLeuGlnLeuAspLysGlyAlaLysAspPheSerProGlyLeu 686
9326 CTGAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9349
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThr 703
9349 CGCT 9399
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaPro 720
9400 GTGCGCAGATCG TCGCTCTCTCTCTCTCTCTCTCTCTCTCT 9443

720 sProThrThrAlaAlaPheAspCysPhePro.....AspThr 732
|||||
9444 CACCTCCACAGGCGCGCCGCAAT...CCCTGATGATGGATACA 9489
733 ThrAlaIAspSerAlaAspSerAlaAsnProPheAlaIAspProLue 749
|||
9490 CACCGCAACGACGCTGATACC.....CCGAGCTCTGGGCGGCCA 9533
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyLeuT 766
|||
9534 GAA.....GATAGCACTCCCGC..... 9552
766 hTrysGlyLeuGlnGlnGlyLysAlaSerAspGlyIleSerLysGly 782
|||
9553GCGCGCGCTCTCGCCTTCGAGCTCCGCGGAC 9588
783 AspThrHis.....GlnAlaSerAlaCysLeuGlyPheGlnI 795
|||
9589 AACGCCACGTCATCTCGAGGAGGCTCCGCGCGCTG.....TC 9629
795 uGlnAspProProGlyGlyLysValAlaSerLeuProGlyAspPheLys 812
|||
9630 GCGCGAGCGCGCACCTCACAGAGGCTGCGAGCGCTCCCGCGCGCT 9679
812 hGlnGlnValGlyGlyValLysGlnGlnAlaGlyLysLeuLeuGlnCys 828
|||
9680 GTCGCGTCTCTGTCGCGAGAGGAGGCGCGCGCGCGCGCGCG 9729
829 ProGln.....ValAlaLysAlaAspArgTrpLeuGlnAs 840
|||
9730 AACGGCTCCGCGCACCACTCTCCGCGAGAGGCTCCGCGCGCTTACGA 9779
840 pSerArgHisCysCysSerThr.....AlaAspPhe GlyAspLeuPro 854
|||
9780 TGGGCTTATCGAGCGCACACCGCGCGCGCTGCGAGCGCGCGCG 9829
855 LeuLeuProProThrSerArgLysGlnAspLeuGlnAlaGlnGlnLys 871
|||
9830 CTCTCTGCGCGCGCGCGCGCGCG..... 9853
871 rSerSerLeuCysGlnLeuLeuGlySerProGlnGlnAlaProGlyMetG 888
|||
9854 AGCTCTCTCCGCGCGCTCTCTCGCGCGCGCGCGCGCGCGCG 9895
888 hAspProLeuSerProLysAlaProLeuIleCysThrLysGlnGlnVal 904
|||
9896CCCCGAGCACGCTCTCGCGCGAGCGGAGGAGCGAG 9931
905 GlnGlnValLeuAspSer.....LysAla.....GlyTr 914
|||
9932 GCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 9981
914 pGlySerProCysHisLeuSerGlyLysValIleLeuLeuGlyProT 931
|||
9982 GCGCT 10031
931 hValGlyThrGlnSerLysValGlnSerTrpPheGlnSerSerLeuSer 947
|||
10032 ATGGAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10060
948 HisMetLysProGlyGlnGlnLysProAsp.....GlyLysAlaVal 962
|||
10061AGTGAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10101
962 oGlyAspSerThrThrSerAspAla..... 970
|||
10102 CCT 10151
971SerLeuAlaGlnLysProAsnLysProAlaVal 981
|||
10152 GGTCT 10201
982 ProGlnAlaProIleAlaLysLysGlnProValProArgGlyLysSer 998

10202 TGTGCGCGCACAGCGAGCGAGATCGCGCGCGCGCTCTCTCTCT 10251
998 uArgSerArgArgValHisArgGlyLeuProGlnAlaGlnAspSerProC 1015
|||
10252 CT 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGln 1028
|||
10287 GCGCAGAGAAAGCGCTCACACCGCTGCGCGCGCGCGCGCGCG 10336
1029 SerCysThrGlyProProGlnGln.....MetGlnGlyAla.. 1041
|||
10337 TCGAGCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10386
1042GlyAlaProGlyArgGlyAlaSerGln 1051
|||
10387 AGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10436
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProArg 1067
|||
10437 CAGAGCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10486
1068ThrProGlyProProGlyLeuThrThrThrProAlaPro..... 1080
|||
10487 TGTTCGCGCGCAAGTCGCGGCTGACTACGCTCTCTCTCTCT 10536
1081ProAspLysLeuGlyG 1086
|||
10537 GAGCGCGTCCAGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCT 10586
1086 LysLysGlnArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
|||
10587 GACGTGCGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10618
1103 ProLysAlaIAspSer.....SerProSerAsnProAla..... 1113
|||
10619 CCGGCGACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10668
1114AlaLeuProValAlaIAspSerSerProMetGlySerLysT 1128
|||
10669 AACCTCCGCGAAGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 10718
1128 hTrysGlnThrAspSerProSer..... 1135
|||
10719 CGATGGCATCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10768
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLysT 1152
|||
10769 TCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10818
1152 hGlnGlnIlePheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168
|||
10839 GG.....CTCATCTGAGAGAGAGAGAGAGAGAGAGAGAG 10849
1169 ProAsnCys.....ArgAl 1173
|||
10850 CCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10899
1173 hThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysVal 1190
|||
10900 GACTGGAAGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10949
1190 rSerSerProGlnLysGln.....GlyArgValSerAlaVal 1203
|||
10950 CACCTACCGCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10999
1204 ArgVal..... 1205
|||
11000 AACGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11049
1206ProLysProGlyAla..... 1210

11050 CTGCTCGGCGCGCGCTGCTTTCGCGCAGCGGGTGGCTTCTTCTTTAC 11099
1211GlySerLysLeuSerAspArg 1217
11100 AGGCGGCTCTCCCTCCGACAGACCCGCTGGCTCGAGGCCATCGCTCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGCACACCATCTCTACCGGCGACCGCTTCTCGAGCTCGCCCTGCAC 11199
1229MetAlaProValProThrLys.....LysArgA 1238
11200 GTGCCCCACCGCTGCTCCGACACCTCGAGACCTCAGCTCGAGGC 11249
1238 snLeuValLeuArgHisLysSerSerSerSerSerSerSerSerSerSer 1254
11250 CCGTCTCGCTCTCCGACAGCAGCAGCTCTCTCTCCAGATCTCGCTCG 11299
1255 Gly..... 1255
11300 GGCCTGAGACGACGACGAGCAGAGCGGCTCTCTTTCATAGCGGACAA 11349
1256GluMetG 1258
11350 GAGGACGCTCTCAGGATGCGCTGAGCTGCGACGCGCAGCGGCTCTCT 11399
1258 LysArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274
11400 CTGCGCGGCGACCCCATCCCTCTCGCGGATCTCCAGGAGGCG..... 11443
1275 LeuProArg.....LysProSe 1280
11444 CTCCTCTGAGTGCCTCCGCTGGACCTCGAGGCTCTACGCAACCTC 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCGAACCTCGGCTGCTACGCGCCGAGTCTCCAGGCTCGGCTCGCTCG 11543
1290 lnSerSer.....His 1293
11544 CTACAGCGCGGACGAGCTCTTTCGAGCGACGAGCTCCCGGAGCGG 11593
1294 ProArgArg...ProProPheLeuProGlnAlaLeuSerAlaAlaPhe 1309
11594 CCGAAAGAGTACCGCGCGCTTTCGCC...TCCACCTCGCTCGCTCGAC 11640
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326
11641 AGCGCCTCGATGCACTGGCTTTGAGGACGAGCAGGAGGAGGCTCGC 11690
1326 eu..... 1326
11691 TCTGCCCTTCCTGAGGCGGAGTCTCGCTCGCTCGCTCGCTCGGTCACCA 11740
1327LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339
11741 CCTTGGCGCTGCTTCCACCGCTCCGAGGCTGAATCTCTCGCTCGATC 11790
1340LeuLysLysPheA 1344
11791 GTCTGGCGGACGCGCGAGTGACCTTTCCTCGGTGCAAGCGCTCGC 11840
1344 laCys..... 1345
11841 CATCGGACGAGTCCGCGCGAGCTCGCAGCCCGGAGCTTCCCAACC 11890
1346LysAlaPro..... 1348
11891 ATGATGGCTCTTCGGCTGCTGAGGAGGAGCTCCAAAGCCCACTCA 11940
1348 1348
11941 CCGCTGCGCGCGCGGCGCTCTTCTCGGACAGGCGCGGCGGAGTCT 11990

1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
11998 CGGCTCGACGCCCGCTCGCTGAGCGGACCTCGCTCGCTTCCGAA 12040
1361 SerAspLysArgGlyLeuLys...GlyAlaLysLysProValG 1376
12041 GCGCTCGACGCGGCTGCTCGCTCGGCTCGCTCGCTCGCTCGCTCG 12085
1376 yValGlnGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392
12086CCTTCATCATGATGACCGGAGGAGCTGCTCCGAGCGGCGGACGA 12131
1393SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404
12132 GGCACCGGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421
12192 GCTTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12231
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438
12232 CACACCGAGAGACGCTCAAGGACCTCG.....CTCAGCGCGGCTCG 12275
1438 LysThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaLysAla 1454
12276 GGGCTCGCGGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12325
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProThrLeuGlyProAl 1471
12326 TCGTGCACATCG.....ACCTCAGGAGGCTCCGACGAGGCTCG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 GTAGGCGCTCGACACGAGAGAGGAGCTCGCTCGCTCGCTCGCTCG 12416
1488 LysAspAsnSerGlyGly..... 1493
12417 ACCCTCATCGGAGTGGCGAACACGCTCGACGAGCGGCTCATCTCC 12466
1494GlyGlyLys.....LysProLysMetGluG 1502
12467 CGCGCAGACACCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12516
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518
12517 GAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12560
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerSerTyrSer 1534
12561 CGCACAGCGCAAGTCCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12610
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
12611 GCGATCTCGTGCACACCTTGGCTCGCTCGCTCGCTCGCTCGCTCGCT 12658
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561
12659TCGAGCGGAGGCGCGGCTCGCTCGCTCGCTCGCTCGCTCG 12692
1561 lnglnGlnValLeuProLeuAspPro..... 1569
12693 CGGTCCAGGTCTCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12742
1570 ...AlaGluProGluIleArgLeuLysTyrIleSerSerCysLysArgLe 1585
12743 TCGGCGCGCTTGGTC..... 12760
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602
12761 CCACGCGCATCGCGGAGCGGCTCGCTCGCTCGCTCGCTCGCTCGCT 12790

[illegible]

8047 CTGGTCGCCAGAGAGCGCCCTGCTCTGCGCCACGAGAGCGGCTCCCA 8096
247 isAspArgProLeuThrAlaSerSerLeuAlaProGlyGlnArgVal 263
8097 TGTCCAGCCCGACAGGCGCTCTCGATC 8125
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280
8125 8125
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 297
8125 8125
297 isHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313
8126TCGGTCTCGATTGCTCA 8143
314 TyrGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 328
8144 TGGCCGTGAGCTTCGCGCGCTTGCAACAGAGCCACCGGCATCAGCTC 8193
328 lArgThrProGlnGlnTyrTyrGlnThrPheSerP 340
8194 CCAGCCACCTCGCTTCGACCATCCCTCTCTCATCGAGTCGCTCTT 8243
340 roSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
8244 CTGCGCGAGCTCGTCGCGCCACGCTCTCGGCACAGAGCTCTCCGTCG... 8290
357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...ProTyr 372
8291AGCCGAGCGCGCGGCTCCGCGGCTTCGCGCGCG 8328
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyIleThrAspH 389
8329 AGCGAGAGCCATCGATCGATCGCGATCGCGATCGCGCTCGCGCGCG 8428
389 isSerHisPheMetProLeuAsnProSer...ProThrAspAlaThr 404
8379 CTGCGCGAGTGCAGCGCTTTGGAGATGCTCTGGCCAGGACCGAGC 8428
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys...As 420
8429 GCGTCGAGCCATTCACAGCGCGA 8454
420 pLysLeuProGlnAsnLeuSerAspLeuSerLeuGlnSerLeuThrA 437
8455 TGGATGCGCGCTGCGCTCTACGACCGCGCGCGCGCGCGCGCGCGAG 8504
437 laLeuThrLeuGlnValGlnAlaGlnLeuSerAsnThrValGlnGlnLeu 453
8505 CTAGCTCCGCGATGCGCGCATGTCGACGAGTGCAGCTCTTCGACCTG 8554
454 LeuSerIysAlaAlaValProGlnLysGlyValIysAsnLeuValSe 470
8555 CTTCTCTTGGCATCAGCCCGCGGAGCGCAACAC 8589
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGlnGlySerG 487
8590 CTCAGCCCCAGCACCGCTGCTCTCGA 8618
487 lYrSerAlaGluProAlaGlyThrProLeuSerGluProPro 501
8619 ATCTGCTGCGAGCGCTTCGAGACCGCGCGCATGTCGCCGCCACCTCA 8668
502SerSerThrProGlnSerThrHI 509
8669 AGGATTCGCCACCGCGCTCTGTCGCGCATCGCGCGCGAGGATACGA 8718
509 sAlaGluProGlnGlnAla 516
8719 TTGCGAGGCGAGCGAGGATTCGCGAGCTTATCGCGCTTATCGCGCTCAAGCA. 8767

516 sPyrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCys 532
8748CAGCGGCTCTTTGCGCGCGCGCTTGGCTACA 8803
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLys 549
8804 GCTTCGCGCTCGAAGGCGCG...CGCTCTCGGTGACACCGCTCTGCTCC 8850
549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSer 566
8851 TCCTGCTCTGCTCGCTCCACCTCGCTCCCAAGCTCTCGACAGGCGA 8900
566 hIysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
890A GTGCAACCTCGCTCGCGCGGCTCTCGGTCTATGGCTCTCCCGGAG 8950
583 SerPheSerLysPheValAla 589
8951 GCTTCGCTCTCTTTCCGCTCGCGCTTGGCGCCGCGCGCTCC 9000
590GlyGluArgAspCysP 595
9001 AAGACCTTCTCGGCAAGCGCGAGCTACGAGCGCGAGAGGCTCAT 9050
595 roArgLeuLeuSerAlaLeuAlaGlnGlnAspLeuAla 608
9051 GCTCTCTGCTCGAGGCTCGGTGAGCGCTCGCGCGCTCGCGGAGGACCGCG 9100
609 SerGluIleLeuGlyLeuGlnGlnAlaIleGlyGlyLysAlaAspLysAl 625
9101 TCCTGCGCTCTGCGCGGCGCGCATCAACGACGACGCGCGCTCGAGC 9150
625 aTrpAlaGluAlaProS 631
9151 GGTATCACGCGCGCCCAAGCGCACTCCAGCAGAGAGTCTCTCGCGCGC 9200
631 erLeuValLysAspSerSerLysProProPheSerLeu 643
9201 CTCTCAGCGCGCGCATCACCGCGCGCTCGAGCTCGAGTCTCGAGTCC 9250
644GluAsnHisSerAlaCysLeuAspSerVa 653
9251 ATGGCAGCGCGACCTCTTGGGAGACCGCATCGAGTGC 9289
653 lAlaLysSerAlaThrProArgProGlyGluProGluAlaLeuProAspS 670
9290AAGCCCTGCGCG...CCGTCTACCGCGCGCGAGCGCG... 9325
670 erLeuGlnLeuAspLysGlyLysAlaLysAspPheSerProGlyLeu 686
9326CTGAAAGCTCTCTCTCTCGCGC 9349
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThrTh 703
9350 CGCTCAGACCAACATCGGCATCTCGAGGCGCGCTCGCGCTCGCGCGC 9399
703 rGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720
9409 GTGCCAGATGTCGCTCTCTCTCGCGCATGAGCGCTCTCGCCCC 9443
720 sProThrThrAlaAlaPheAspCysPheProAspThr 732
9444 QACCTCCACAGCGCGCGCGCAT...CCCTTGATGATTGGGATACA 9489
733 ThrAlaAlaSerSerAlaAspSerAlaAsnProPheAlaThrProGluG 749
9490 CTCGCCATCGAGCTGTTGATACCCCGAGGTCTGGCGCGCGCA 9533
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeu 766
9534 CGAAGATAGAGTCCCGCG... 9552

766 hrlysglyleucluglyllysAlaseraspolyIleuSerlysgly 782
 9553
 783 AspThrHis.....GluAlaserAlaCysleucluglyl 795
 9589 AACGCCACGATCATCTCGAGAGAGGCTCCGCGCCCTG.....TC 9629
 795 uGluAspProGlyGlyLysValAlaSerLeuProGlyAspPheIysG 812
 9630 GGGCGAGCCCGCACCTCACAGACGCTCCGACCGCTCCCGCGCGCT 9679
 812 lngluGluValGlyLysValGlyLysGluAlaGlyLysleuclugly 828
 9680 GTGCCGTGCTCTCTGCGCCAGAGAGGCGCCGCTCCGCGCCAGCG 9729
 829 ProGlu.....ValAlaLysAlaAspArgTyrleuGluAs 840
 9730 AACGGCTCCGCGACACCTCTCCCTCCACAGACCTCCCTTATCGA 9779
 840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854
 9780 TGTGGCTATTTCGAGGCGACACCGCGCCACTTCGAGCGCGCGCG 9829
 855 LeuLeuProThrThrSerArgLysGluAspLeuAlaGlyLysGly 871
 9830 CTCTCCTGCGCCCGGACCGGAG..... 9853
 871 rSerSerLeuGlyGlyLeuLeuGlySerProGlyGluArgProGlyMetG 888
 9854 .AGCTCTCTCCGCGCTCGACTCTCCGCGCCAGACAAAGCCG..... 9895
 888 lAspProLeuSerProLysAlaProLeuIleCysThrLysGlyLysVal 904
 9886CCCCGAGCACCGTTCTCGCGCCGAGCGAGCGAACCCACG 9931
 905 GluGluValleuAspSer.....LysAla.....GlyTyr 914
 9932 GCAAGGTGCTCTGCTCTTCGAGCAAGGCTCGCAGTGGAGAGGATG 9881
 914 pGlySerProCysHisLeuSerGlyLysValIleLeuLeuGlyProT 931
 9982 GCCCTCTCCCTGCTCGACTCTCCGCGGCTTCGCGCTCTACTCTGAAGC 10031
 931 hrValGlyThrGluSerLysValGlnSerTyrPheGluSerSerLeuSer 947
 10032 ATGGAGGCGCGCGCTCGCTCCACGCTG..... 10060
 948 HisMetLysProGlyGlyGlyLysProAsp.....GlyGluArgAlaPr 962
 10061AGTGGAGCTCTCGCGCTCTCGCGCCGCGACGAGGCGCC 10101
 962 oGlyAspSerThrThrSerAspAla..... 970
 10102 CCTTCCCTCGACCGCGCTCGACGCTCGACGCCCTCTCTTCCGCTAT 10151
 971SerLeuAlaGlnLysProAsnLysProAlaVal 981
 10152 GGTCTCCCTGCGCGCCCTCTGCGCTCGCTCGAGCCCGCGCGCG 10201
 982 ProGluAlaProIleAlaLysLysGluProValProArgLysSerLe 998
 10202 TCGTCGCGCCACAGCGAGGCGAGATCGCGCGCTTCGTCGAGGCGCT 10251
 998 uArgSerArgArgValHisArgLysLeuProGluAlaGluAspSerProC 1015
 10252 CTCTCCCTCGAGAGACGCGCGCGCATCGCG.....CCCT 10286
 1015 ys.....ArgAlaProValleuProLysAspLeuLeuProGlu 1028
 10287 GCGCAGAGAAAGCGCTCACCGCTCGCGCGCACGCGCGAGTGGCGCG 10336
 1029 serCysThrGlyProProGlnGlyGln.....MetGluGlyAla.. 1041

10337 TCAGACTCGCGCGCTCCGACCTCCAGACCTACCTCCCTCCGCGCGAG 10386
 1042GlyAlaProGlyArgGlyAlaSerGlu 1051
 10387 AGGCTCTCCACCGCGCGCTCAACAGCCCGAGGCTACCTCTGATCCG 10436
 1051 .LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProArg 1067
 10437 CGAGCCCGCGCGCGCTCGAGCGCTCTCGAGCTCTCCACCGCACAG 10486
 1068ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080
 10487 TGTTCGCCCGCAAGTCCGCGTGAAGTACGCTCCGCTCCGCGCAAGT 10536
 1081ProAspLysLeuGly 1086
 10537 GACGCCGTCCAGACAGCTCGCGCGAGGTCTACCAAGATCGCTCTCG 10586
 1086 LysGluArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
 10587 GAGCTCGAGCTCCCTCTTATT.....CGACCTCA 10618
 1103 ProLysAlaAlaSer...SerProSerAspProAla..... 1113
 10619 CCGGACACAGGCTCGAGGCTCCGAGCTCGAGCGCGCTGATCTGATCGA 10668
 1114AlaLeuProValAlaSerAspSerSerProMetGlySerLysT 1128
 10669 AACCTCGGCAACCGCTCTGCTCTCGAGCGACGACGACGAGGCTCTCGA 10718
 1128 hrLysGluThrAspSerProSer..... 1135
 10719 CGAGGCGACATCGCTCTCTCGAGGCTCGAGCTCCACCGACGCTCTCG 10768
 1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLysT 1152
 10769 TCGCCCTCGCGAGACTCGAGCGCTCGAGCTCGAGTCCGCTCTCTCTC 10818
 1152 hrGlnGluIlePheHisSerLysArgArgTyrProSerGlyLysGlyLeu 1168
 10819 GG.....CTCATTCAGACGAGAGAGGCCACTCG..... 10849
 1169 ProAsnGly.....ArgAl 1173
 10850 CCGCGCTGCTCTCTCTCGAGGCGAGCTCTACCGAGGCTCTCGCTC 10899
 1173 aThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190
 10900 GACTGGAAGGACTTCTCGCGCGCTCTCGAGCTCCGCAAGGCTCTCCCTCC 10949
 1190 .aSerSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203
 10950 CAGCTACCCCTTCAGAGAGAGAGGAGTTCGCTCGAGTCTCCAGAGAG 10999
 1204 ArgVal..... 1205
 11000 AACGCTTCGAGAGTCCCTCCGAGGCTGACTCGCGCGACCAATCCCG 11049
 1206ProLysProGlyAla..... 1210
 11050 CTGCTCGCGCGCGCTCGCTCTCTCGCGCGAGGAGGCTCTCTCTTAC 11099
 1211GlySerLysLeuSerAspArg 1217
 11100 AGGCGGCTCTCTCTCGAGAGACCGCGGCTCGAAGGATATGCCGTCT 11149
 1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
 11150 TCGGACACACCATCTACCGGAGCACCGGCTTTCGAGCTCGCGCTCGAC 11199
 1229MetAlaProValProThrLys.....LysArgA 1238

11200 GTCCGCCACCGCTGGCTCGACACCGTCGAGAGCTACGCTGAGGC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAsnAlaSerAlaMet 1254
11250 CCTCTCGCTCTCCATCGACGACACCGCTCTCTCCATCGCTCGCG 11299
1255 Gly 1255
11300 GCGCCGTCGACGACGACGACGACGACGACGACGACGACGACGAC 11349
1256 1256
11350 GAGGACGCGCTTCAGATGCGCCCTGAGCTCGCCACGCGAGCGCTCTCT 11399
1258 LysArgArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274
11400 CTCGCGCGGACGACGACGACGACGACGACGACGACGACGACGAC 11443
1275 LeuProArg 1280
11444 CTCCTCGAGTGCCATCCCGGTGAGCTCGAAGGCTCTACGCAACCTTC 11493
1280 rPro 1290
11494 GCAACCTCGGGCTTCGCTACGCGCGGAGCTTCAGGCGCTTCGCTCCCT 11543
1290 InSerSer 1293
11544 CTCACGCGCGCGCGAGCTCTTTGCGAAGCGCAAGCTCCCGGAAGCGG 11593
1294 ProArgArg 1309
11594 CGCAAGAGATGCGCGCGGCTTTGCC...TCCACCTCGCGCTCGTAC 11640
1309 ednGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326
11641 AGCGCCCTCGATGCTGCTTCGAGGACGACGACGACGACGACGCTCGC 11690
1326 eu 1326
11691 TCTGCCCTCTCGTGGAGCGGAGTCTCGCTCGCTCGCTCGCTCGCACC 11740
1327 1327
11741 CCTTGGCGTCTCGCTTCCACCTCGCTCCAGAGGTGAATCTCGCTCGATC 11790
1340 1340
11791 GTCTGGCGGACGCGCGAGGTGACCTCTTGCCTCGGTGCAAGCGCTCGC 11840
1344 laCys 1345
11841 CATGCGGACGCTCGCGCGCGAGCTCGCGACGCTCGCGACGCGCTCGCAC 11890
1346 1346
11891 ATGATCGCTCTCGCGGTGCGACTGAGCGAGCTTCCAAAAGCCCACTTCA 11940
1348 1348
11941 CCGCTGCGCGCGCGCGGCTCTTTCGCGACAGGCGCGCGACGATCT 11990
1349 1349
11991 CGCGCTCGAGCGCGCTCGCGGTGACGCGAGCTCGCTCGCTCGCTCGGAA 12040
1361 SerAspLysAspArgGlyLeuLys 1376
12041 CGCGCTCGACGAGGCGCTTCGCTCGCGCGCTCGCTCGCTCGCTCGC 12085
1376 yValGluGluGlyLeuValAlaValGlyThrGlyGluLysLeuProThr 1392
12086 12086
12131 12131

1393 1393
12432 GGCACGCGCTCGCTCGCTCTTCAGAGCTGGCTCGCTCGCTCGCTCG 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421
12182 GCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12231
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGly 1438
12232 CACACGAGACGACGCTCAAGACCTCG...CTCAGCGCGCGCTCG 12275
1438 LysThrAlaLeuAlaProLysLysArgSerArgLysGlyValGluAlaGlyAla 1454
12276 GGGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12325
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrLeuGlyProAl 1471
12326 TCGTGCACATCG...ACCTCAGGAGGCTTCACGAGGCGCTG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 CTAGCGCGCTCGACACGAGGACGCTCGCTCGCTCGCTCGCTCG 12416
1488 LuAspAsnSerGlyGly 1493
12417 ACCCTCATCCGAGGTGGCGACACGCTCGACGCGCGCTCATCC 12466
1494 1494
12467 GCGCGACGACGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12516
1502 uLeu 1518
12517 GAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12560
1518 LntThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerTyrSer 1534
12561 CGCACGCGCGAGTTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12610
1535 LysArgLys 1544
12611 GCGATGCTCGACACCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12658
1544 aLysAsnThrThrSerSerProCysLysGlyValGluAlaLysArgArg 1561
12659 12659
1561 LngLingLingValLeuProLeuAspPro 1569
12693 GGTTCAGGTGCTCCGATACACCTGAGCGCGCTGATGGGGTCT 12742
1570 1570
12743 TCGCGCGAGCTTTGCTC... 12760
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGlu 1602
12761 12761
1602 ysArgAspAlaPheThrThrileCysThrValValAsnSerProGlyAsp 1618
12791 12791
1619 AlaProLysProHisArgLysProSerSer 1628
12813 TCGCGCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12864
1629 1629
12865 CTCGCTGATCTCGGCGATCTGAACCCCAATCAACGCTGCTCACTCAT 12914

alignment_scores:	
Quality:	493.00
Ratio:	0.593
Percent Similarity:	39.666
Percent Identity:	22.291
alignment_block:	

Align seg 1/1 to: US-08-458-076A-6 from: 1 to: 28958

62	GLYALALVTHrProserGlyTHrAlaAlaValAla.....Al	75
63		
64		
65		
66		
67		
68		
69	GGCGCGAGAACAGAGCGCTACGCGCGCGCAAGCGCTCTCGAGCG	7658
70	asPLySYrNHsArGlySerIyAlaLeuProthrGln.....	89
71		
72		
73		
74		
75	CGTGGCGAGCATCGCGCGAGCGCTGGATACAGGAGCGCGGTGGCT	7708
76		
77		
78		
79	GGGCGCGTGGCGCGCGCGCGCGAGCGACATAGCGGCGAGCGCAC	7758
80		
81		
82		
83		
84		
85		
86		
87		
88		
89		
90	..GLyeuGlnIlyArgProIlePheProGlyTYrGlyValGlnAspse	105
91		
92		
93		
94		
95		
96		
97		
98		
99		
100		
101		
102		
103		
104		
105		
106		
107		
108		
109		
110		
111		
112		
113		
114		
115		
116		
117		
118		
119		
120		
121		
122		
123		
124		
125		
126		
127		
128		
129		
130		
131		
132		
133		
134		
135		
136		
137		
138		
139		
140		
141		
142		
143		
144		
145		
146		
147		
148		
149		
150		
151		
152		
153		
154		
155		
156		
157		
158		
159		
160		
161		
162		
163		
164		
165		
166		
167		
168		
169		
170		
171		
172		
173		
174		
175		
176		
177		
178		
179		
180		
181		
182		
183		
184		
185		
186		
187		
188		
189		
190		
191		
192		
193		
194		
195		
196		
197		
198		
199		
200		
201		
202		
203		
204		
205		
206		
207		
208		
209		

8125 8125
297 I sHisalaglndlnthrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313
8126 TGGGTCTGATTGCTCA 8143
314 TyrGlycIndlnGlynglyTyrCysGln..... ProAspAlaAla 328
8144 TGGCCGTCGAGTTCGCGGGCTTCGACAGCGCCACCGCAGCTCAAGTC 8193
328 largThrProGluGln..... TyrTyrGlnThrPheSerP 340
8194 CGGCCACCTCGCTTCGACCATCCTCTCTCATGAGTCGCTCTT 8243
340 toSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
8244 CTTGCGGACTCGCTCGCCACCGCTCGGCACGAGGCTCTCGTCG... 8290
357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe... ProTyr 372
8291 AGCCGACGCGCGCGCTCGCGGCTTCGCGCGCG 8328
372 rSerGlnInProLeuSerThrGlyAlaPheProAlaGlyLeuThrAspH 389
8329 AGCGACGAGCCATCGCATCGTCGATGCGCTCGCGCGCG 8378
389 isSerHisPheMetProLeuAsnProSer... ProThrAspAlaThr 404
8379 CGTCGGGATGTCGAGCTCTTGGGAGTTCCTGCGCCAGGACGCGAG 8428
405 SerSerValAspThrGlnAlaGlyAsnCysLysProLeuGlnLys... As 420
8429 CGGTCG..... AGCCCATTCACAGGCCGCA 8454
420 pLysLeuProGluAsnLeuSerAspLeuSerLeuGlnSerLeuThrA 437
8455 TGGATCCGCTCGCTCTAGACCCCGCGACCGACCAAGCAAGAG 8504
437 IaLeuThrLeuGlnValGluAsnIleSerAspThrValGlnGlnLeu 453
8505 CTAGCTCGGCGATGCGGCATGCTGACACGATGCTGCTTGCACCTG 8554
454 LeuSerLysAlaAlaValProGlnLysLysGlyValLysAsnLeuVal 470
8555 CTTCTCTTGGCATCAGCCCGCGGAGGCAAC..... AC 8589
470 rArThrProGluGlnHisLysSerGlnHisCysSerProGluGlySer 487
8590 CTGACCCCTC..... AGCACCGCTGCTCC..... TCGA 8618
487 lYrTyrSerAlaGluProAlaGlyThrProLeuSerGluProPro..... 501
8619 ATCTGCTTGGAGGCTTCGAGACGCGCGCATGCTCCCGCCACCTCA 8668
502 SerSerThrProGlnSerThrH 509
8669 AGGATTCCTCCACCGGCTCTCTGTCGCAATGCGCGCAAGCAATACGA 8718
509 sAlaGlu..... ProGlnGluAla 516
8719 TTGGAGAGCGGACCGAGATTCGACGCTTATGCTTCACAGGCA. 8767
516 sPyrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCys 532
8768 CGCGCGGCTCTTTCGCGCGGCGCTTGCCTACA 8803
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerHisSerLysAla 549
8804 CGCTCGGCTGAGAGGCGCG... CGCTCTGCTGACACCGCTGCTCC 8850
549 sProLysValSerThrCysSerValThrSerProAspAspMetSer 566
8851 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 9679

566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
8901 FTGCAACCTGGCCCTCGCGCGGCTCTCGCTCATGCTCCCGCGAG 8950
583 SerPheSerLysPheValAla..... GlyGluArgAspCysP 595
8951 GCTTCGCTCTCTTTCGCGCTCGCGCTTGGCGCGCGAGGCTCAT 9050
590 GlyGluArgAspCysP 595
9001 AGACCTTCTCGGCCAACCGCGAGCTACGACGCGGAGAGGCTCAT 9050
595 roArgLeuLeuSerAlaLeu..... AlaGlnGluAspLeuAla 608
9051 GTCCTTTCGCTCGAGGCTCGCTGACGCTCGCGCGCGAGGACACCG 9100
609 SerGluLeuLeuGlyLeuGlnGluAlaIleGlyGluLysAlaAspLysAl 625
9101 TCTCGGCTCTGCTCGCGGACCGCTCATACCGAGCGGCGCTGAGC 9150
625 a..... TrpAlaGluAlaProS 631
9151 GGTATCACGCGCCCAACGCGACCTCCACAGAGAGTCTCGCGCGCG 9200
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643
9201 GCTCCAGAGCGCGCATCACCGCGCGAGCTCGACGCTCGTCTGAGTGC 9250
644 GluAsnHisSerAlaCysLeuAspSerVa 653
9251 ATGGCACCGGACCTCTTGGAGACCCCATCGAGGTGC..... 9289
653 lAlaLysSerAlaTrpProArgProGlyGluProGluAlaLeuProAspS 670
9290 AGCCCTGCGCG... CGCTCTAGCGCGACGCGAGACCG... 9325
670 erLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerProGlyLeu 686
9326 CTGAAAGCTCTCTCTCTCGCGG 9349
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThrTh 703
9350 CGCTCAAGACACATCGGCTATCGAGCGCTCGCGCTCGCGGCG 9399
703 rClyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720
9400 GTGCGCAGATG..... TCGCTCTCTCGCGCATGACGCTCGCGG 9443
720 spProThrAlaAlaPheAspCysPhePro..... AspThr 732
9444 CACCTCCACAGCGCGCGCGCAT..... CCTTGTATGATTGGGATACA 9489
733 ThrAlaLysSerAlaAspSerAlaAsnPropheAlaTrpProGluG 749
9490 CTCGCCATCGAGCTGTGATAC..... CGAGGTCTTGGCGCGGCA 9533
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuT 766
9534 CGAA..... GATAGAGTCCCGC..... 9552
766 hrLysGlyLeuGlnGlyGlyLysAlaSerAspGlyIleSerLysGly 782
9553 GCGCGGCTCTCGCGCTTGGAGCTCTCGCGGAC 9588
783 AspThrHis..... GluAlaSerAlaCysLeuGlyPheGlnG 795
9589 AACCCACATCATCTCGAGGAGGCTCCCGCGGCTG..... TC 9629
795 uGluAspProGlyGluLysValAlaSerLeuProGlyAspPheLysG 812
9630 GGGCGAGCGCGCACCTCACAGAGGCTCGGCGGCTCCCGCGGCT 9679

[illegible][illegible]

1675 alserlysa...LeuserTherCys 1683
 13053 TGACCTGGGCTTCGAGCAGCACTCTGTC 13081

seq.name: /cgn2_6/ptodata/2/lna/5A.COMB.seq:US-08-764-233A-4

seq_documentation_block:

Sequence 4, Application US/08764233A
 Patent No. 5716849

GENERAL INFORMATION:

APPLICANT: Ligon, James M.

APPLICANT: Schnupp, Thomas

APPLICANT: Beck, James J.

APPLICANT: Hill, Dwight S.

APPLICANT: Neff, Suzanne

APPLICANT: Ryals, John A.

TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,233A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/729,214

FILING DATE: 09-OCT-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/258,261

FILING DATE: 08-JUN-1994

ATTORNEY/AGENT INFORMATION:

NAME: Weigs, J. Timothy

REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: 1506/CIP6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 541-8587

TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Sorangium cellulosum

IMMEDIATE SOURCE:

CLONE: p98/1

US-08-764-233A-4

alignment_scores:

Quality: 493.00

Ratio: 0.593

Percent Similarity: 39.666

Percent Identity: 22.291

alignment_block:

US-09-508-821b-6 x US-08-764-233A-4

Align seg 1/1 to: US-08-764-233A-4 from: 1 to: 28958

62 GYALAGLYThrProSerGlyThrAlaAlaAlaValAla.....Al 75

||||:||||| :||| ||||| |||

7609 GCGCGCGAGACAGACAGCGGCTACGCGCGCGAAGCGCTCTCGACGC 7658

75 asprlystThrHisArgGlySerLysAlaLeuProThrGln..... 89

7659 CATTCCGAGATCGCGCGCTGAGTGGATGGCCCTCGCTGCGCGTGGCT 7708

90 ..GlyLeuGlnGly ArgProAlaPheProGlyTyrGlyValGlnAsp 105

7709 GGGGCGCGTGGGCGCGCGCGCGCGACAGCGACAGCGCGCGCGCGAC 7758

105 rser.....ProTyrProGlnArgTyrAlaGlyGln 116

7759 CTCGACAGCGCGCGCTCGCGGATGGCCCTCGCTGCGCGTGGCGG 7808

116 LserLeuGlnAlaTyrGlyAlaProGlnProProGlnPro... 131

7809 GCTCGCGCTGCTGAGACAGCGACAGCGACAGCGACAGCGACGAC 7858

132GlnProLeu...ProAlaGln 137

7859 TCGACTGGCGCGCTTGGCGCTTGGTACGCGCGCGCTGGCGCGCG 7906

137 yValAlaLysTyrAspGlnAsnLeuMetLysTyrAlaValProPhe 154

7907CGCTCTGCGCGATTTGCCCGA 7928

154 asrlystThrAlaGlnGlnGlyAlaGlnValProPheArgThrHis 170

7929 GGGCGCGCGCGCTCGACAGCGCGAGGCG..... 7960

171 LeuHisValGlnGlnProProProGlnGlnProLeuAlaTyrPro 187

7961CGCTCTGCGAGCATGGCGCGCGCG..... 7984

187 sleuGlnArgGlnLysLeuGlnAsnAspLeuAspLeuProPhe 204

7984 7984

204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSer 220

7985CGAGCGCTCTCG 7996

221 ThrTyrSerSerSerValGlnGlyLysGlyAlaHisSerTyr 237

7997 ACAAGCTCGAGCGCGCTCGAGAGCGAGCGAGCGAGCTGCTGCTCG 8046

237 sSerCys.....ThrAlaProThrAlaGlnPro 247

8047 CTGGAGCGCGACAGAGCGCGCTCTCGTCCGCGACAGAGCGCGCTCG 8096

247 IsAspArgProLeuThrAlaSerSerSerLeuAlaProGlyGlnTyr 263

8097 TGTGAGCGCGACAGAGCGCGCTCTCGATC..... 8125

264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGln 280

8125 8125

280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeuGlnSerArg 297

8125 8125

297 IsHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313

8126TCGGTCTCGATTCGCTCA 8143

314 TyrGlyGlnGlnGlnGlnGlnTyrCysGln.....ProAspAlaAla 328

8144 TGGCGGTGAGCTTGGCGCGCTTGCACAGCGCGCGCGCATCAAGTC 8193

328 IsArgThrProGlnGln.....TyrTyrGlnThrPheSerP 340

8194 CGGCGCGCGCTTGCACAGCGCGCTTGCACAGCGCGCGCGCTTGCAC 8243

1290 lnsrSer.....His 1293
: : : : :
11544 CTACAGCGGGGAGAGAGCTTTTCGCGAGGCAAGCTCCCGGAGCGG 11593
1294 ProArgArg...ProProPheLeuProGluAlaArgLeuSerAlaAlaph 1309
: : : : :
11594 CCGAAAGAGTCCGCGCGGTTGCCC...TCCACCTCGGCTCGAC 11640
1309 eGnglyAlaMetLysThrLysValLeuProProArgLysGlyArgLyl 1326
: : : : :
11641 AGCGCCCTGCATGCACTGGCCCTTTGAGGACGACGAGAGGAGGCTCGC 11690
1326 eu..... 1326
11691 TCTGCCCTTCCTCGGAGGAGAGTCTGCTCGCTCCGCTCGGCTGCCACCA 11740
1327LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339
: : : : :
11741 CCTTGGCGGCGCTTCCACCTCCCAAGGCTGAAATCTTCGCTCTCGATC 11790
1340LeuLysLysPheA 1344
: : : : :
11791 GTCTGGCGGACGCGGAGGTGACCTCTTGGCTCGGTCAGAGGCTCGC 11840
1344 lacys..... 1345
: : : : :
11841 CATGGGAGACAGTCCGCGCGAGCTCCGACCCCGGAGCTTCCCAAC 11890
1346LysAlaPro..... 1348
: : : : :
11891 ATGATGCCCTTTCGCGTCTGAGGAGCTCCGAGGCTCCAAAGCCCACTCA 11940
1348 1348
11941 CGCCTGCGCGCGAGCGGCTCTCTCGGACAGGCGGCGCAGATCT 11990
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
: : : : :
11991 CGCGTCGACGCCCGCTCGCCGCTAGCGGACCTCGCTGCCCTCCGAA 12040
1361 SerAspLysAspArgGlyLeuLys...GlyAlaGlyLysProValGl 1376
: : : : :
12041 GCGCCTCGACAGGCGCTTGGCTCCCGGCTCTGCTGCTGCCC..... 12085
1376 yValGluGluLeuValAsnValGlyThrGlyLysLeuProThr. 1392
: : : : :
12086CCTTCATCATGACGCGGAGGAGCTGCTCCGAGGCGCCACGA 12131
1393SerGlyAlaAspProLeuLysArg.....AsnProThrAsn 1404
: : : : :
12132 GGCACCGCGCTCGCACTCGCCCTTTGCAAGCTGCTCGCGGACGAA 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421
: : : : :
12182 GCGTGGCTCTCGCGCTTCGCTTCGCTACCGGAGCGGCTGCGCAC 12231
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGly 1438
: : : : :
12232 CACACGGAAGACGCTCAAGGACCTCG.....CTCACGCGCGCTCTG 12275
1438 lYthAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454
: : : : :
12276 GGGCTCGCGGCTCGCGCAAGTGAAGACCCAGACCTCCGCTCTTC 12325
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyrLeuGlyProAl 1471
: : : : :
12326 TCTCGACATCG.....ACCTACCGAGGCTCCCGAGGCGCCCTG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
: : : : :
12367 CTAGGCGGCTCGACACAGGAGACGCGAGCTCGCTCCGCAAGGGAA 12416

1488 lueAspAsnSerGlyGly..... 1493
: : : : :
12417 ACCCTCATCCGAGGTGGCAACACCGCTCGACGAGCGCTCATCC 12466
1494GlyGlyLys.....LysProLysMetGluGl 1502
12467 CGCGGCAACACCGAGTGGCGCTCCATATTCGCAAAAGGACCTTC 12516
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518
: : : : :
12517 GACGCGCTCGCCCTCGTCGACGCCCG.....AGCCCGAGGCGCCCT 12560
1518 lnhThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerTyrSer 1534
: : : : :
12561 CGCACAGCCCAAGTCCGCATCGCGTCGACGCGGAGGCTCAACTTC 12610
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
: : : : :
12611 GCGATGCTTCGACACCTTGGCATATATCCGGGCGAGCGCGCGCGC... 12658
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArg 1561
: : : : :
12659TCGGAGCGGAAGCGCGGCGATCGTACTGAAGT 12692
1561 lnhGlnValLeuProLeuAspPro..... 1569
: : : : :
12693 CGGTCCAGGTGCTCCCGATACACCGTAGCGGCGGTATGGGGGTCT 12742
1560 ..AlaGluProGluIleArgLeuLysTyrIleSerSerCysLysArgLe 1585
: : : : :
12743 TCGGCGGACCTTTGGTC..... 12760
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602
: : : : :
12761 ..CCAGGCCATCGCGAGCGCGCGCATGATCT..... 12790
1602 ysArgAspAlaPheThrIleCysThrValValAsnSerProGlyAsp 1618
: : : : :
12791GCCCATCCCGCCAGCGCTGCTCT 12814
1619 AlaProLysProHisArgLysProSer..... 1628
: : : : :
12815 TCGCCCAAGCCCGAGCGCTCCCATCATCTATCTACCGCTACTATGA 12864
1629SerAlaSerSerSer..... 1633
: : : : :
12865 CTGCTGATCTCGGCATCTGAAACCAATCAACGCTGCTCATCCATGC 12914
1634SerSerSerSerSerSerLeuAspAlaAlaGly 1645
: : : : :
12915 GCGCGCGCGCGCTCGGAGCGCGCGCTTCTAGCTCGACGCGACCTCG 12964
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPro 1659
: : : : :
12965 CGCGGAGGTCTTCCACCGCGCGCTCCAGGAGAGTGGAGCGCTCTCCG 13014
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675
: : : : :
13015 GCGCTCG.....GTTCCGAGGATGGGACCTCGCGCTCTCAG 13052
1675 alserLysAla...LeuSerThrSerCys 1683
: : : : :
13053 TGACCTGGGCTTCGAGGACGACTTCTCTGC 13081
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-457-335A-6
seq_documentation_block:
; Sequence 6, Application US/08457335A
; Patent No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph

[illegible][illegible]

```

10769 TCAGCCCTCCGAGAACCTCGAGAGCCCTCACAGCTGCATCCGCTGCTGTC 10818
1152 hrcInsl1LePheNslasrLysAaYgAaYrProSerGlnLysAaGLeu 1168
10819 GG.....CTCCATTCGACGAAGAAGAAGCCAACTCG..... 10849
1159 ProAsnCs.....:|||||:|||||:|||||:.....ArgAl 1173
11580 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1189
10850 :CCGCGCTCTCTCTCTCTGAGGAGAGACTCTTACCCGAGAGCCCTGAGCTC 10899
1173 athrLysLysLeuLysAspLysSerHisLeuProAlaThrPheLysValS 1190
10900 GACTGGAAAGACTTCTTCGAGCCCACTCAAGCTCCCGGAGAGTTCCTCCCTCC 10949
1190 eArSerSerProGlnLysGln.....GlyAaYALserGlnAaYALa 1203
10950 CAGCTCAACCCCTTCCAGAGAAAGAGTCTGAGCTCGACGCTCCAGAGAGC 10999
1204 ArgVal.....:|||||:|||||:|||||:|||||:|||||: 1205
11000 AAGCTTCCGACGTGCGCTCCGAGAGCTACACTCCGCGGACCAATCCG 11049
1206 .....ProLysProGlnYALa.....:|||||:|||||:|||||: 1210
11050 CTGCTGAGGAGCGCGCGCGCTCTCGAGACCGAGCGGTTCCTTCCTTAC 11099
1211 .....:|||||:|||||:|||||:|||||:|||||:|||||: 1217
11100 AGGCGGAGCTCTCCCTCCGACAAAGACCCGCTGGCTCGAAAGCCATGCCCT 11149
1218 ProLeuHisAlaLeuLysAaYrGlySerAlaPhe.....:|||||: 1228
11150 TCGGCAACCCCATCTTACCGGAGCAACGGAGTTCTTCGAGCTCCGCTGAC 11199
1229 .....MetAlaProValProThrLys.....LysAaYAL 1238
11200 GTGCGCCACCGCGTGGCTCGACACCTCGCAAGCAAGTCAGCTGAGAGCC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254
11250 CCGCTGCGCTCTCCCATCGAGGACACGCTCTCTCGAGATCTCCGTCG 11299
1255 GLy.....:|||||:|||||:|||||:|||||:|||||:|||||: 1255
11300 GGGCCGTGACAGACGACGAGAGGCGCTCTTTCATAGCCGACA 11349
1256 .....GluMetG 1258
11350 GAGGAGCGCTTCAAGATGAGCCCTCGAATGCCACGCCAGCGGCTCTCT 11399
1258 LysAaYrGlyArgLysLeuArgValProProProSerSerArgLysLeu 1274
11400 CTGCGCGGAGACCCCATCCCTCTCCGCGCATCTCCAGAGATGAGC..... 11449
1275 LeuProArg.....:|||||:|||||:|||||:|||||:|||||: 1280
11444 CTCCTCTGAGTGCATCCCGATGAGACTCGAAGGCTCTACGACCACTTC 11499
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCCAAGCTGAGGCTTGCCTACGCGCCCGAGTTCAGAGGCTTCGCTCCGT 11549
1290 InSer.....:|||||:|||||:|||||:|||||:|||||: 1293
11544 CTACAAAGCGGCGAGAGCTTCTTCCGCAACCAAGCTCCGAGAAAGCG 11599
1294 ProArgArg.....ProProPheLeuProGlnAlaArgLysSerAlaAlaP 1309
11594 CCGAAAGAGATGCGCCCGGTTGCCC.....TCCAACTCGCGGTGCAC 11649
1309 eGlnGlyAlaMetLysThrLysValLeuProArgAlaGlyArgGlyL 1326
11644 AGCCGCTGCAATGCACTGCGCTTTCAGAGACAGACAGAGAGGACGCTGC 11699

```

1326 eu..... 1326
11691 TCTGCCCTTCGTGGAGCGAGTCTCGCTGGCGCTCCGTGGGTGCCACCA 11740
1327LysLeuGluAlaIleValGlnLysIleThrSerProSer..... 1339
11741 CTTGGCGGTGGCTTCCACCTCCCAAGGAGTGAATCTCGCTTCGATC 11790
1340LeuLysIlePheA 1344
11791 GTCTGGCGGAGCGCGAGTGAACCTCTTGGCTCGGTGCAAGGCTCGC 11840
1344 lacys..... 1345
11841 CATGGGAGACGTCTCGCGCGAGCTCCGAGCCCGGAGCTCCACCC 11890
1346LysAlaPro..... 1348
11891 ATGATGCGCTCTTCGGGTGACTGGAGGAGCTCCAAAGCCCACTTCA 11940
1348 1348
11941 CGCGCTGCCCGCGAGCGCGCTCTTCTGGCAGAGCGCGCAGCATCT 11990
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
11991 CGCGCTCACCGCCGCTCGCGCGTAGCGGACCTCGCTGCCCTCCGAA 12040
1361 SerAspLysAspArgLysLeuLys...GlyAlaGlySerProValG 1376
12041 GCGCCTCGACGAGGCGTCTGCTCCCGGCTCGCTCGCTCGCC..... 12085
1376 yValGluGluLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392
12086 ...CCTTCATCATGAGCGAGCGAGCTCGTCCGAGCGCGCCACGA 12131
1393SerGlyAlaAspProLeuLysArg.....AsnProThrAsn 1404
12132 GCGCACCGGCTCGCACTCGCTCTTCGAGCGCTCGCTCGCGCAGCAAC 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421
12182 GCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12231
1421 pCysPheLysThrCluAlaPheThrSerProGluAlaLeuGlnProGlyG 1438
12232 CACACGAGAGCAGCTCAAGGACCTCG.....CTCAGCGCGCGCTCG 12275
1438 lyThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454
12276 GGGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 12325
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyLeuGlyProAl 1471
12326 TGTCTGACATCG.....ACCTCAGAGGCGCTCCAGCAGGCGCTG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 CTAGCGGCTCGCTCGACAGGAGGAGCGCTCGCTCGCTCGCTCGCTCG 12416
1488 luAspAsnSerGlyGly..... 1493
12417 ACCCTCATTCGAGGTGGCGAGCAGCAGCTCGAGCGAGCGCTCATCC 12466
1494GlyGlyLys.....LysProLysMetGluG 1502
12467 CGCGCAGCAGCAGCAGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12516
1502 uLeu...GlyProLaserGlnProProGluGlyArgProCysGlnProG 1518
12517 GAGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 12560

1538 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTyrrSerSerTyrrSer 1534
12561 CGCACGCGCAAGTCCGATCCCGCTCGACGCGCGAGGCTCAACTTCC 12610
1585 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
12631 CGCATCTCGCTCGACACCTTGGCATATCCGGGCGAGCGCGCGCGC... 12658
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561
12659TCGGAGCGAAGCGCGCGCATCGTACTGAAGT 12692
1563 lnGlnGlnValLeuProLeuAspPro..... 1569
12693 CGTCCAGGTGCTCCCGATACACCTAGGCGACCGGGGTGATGGGGGTCT 12742
1570AlaGlnProGluLeuArgLeuLysTyrrIleSerSerCysLysArgLe 1585
12743 TCGCGCAGCCTTTGGTC..... 12760
1585 uArgSerAspSerArgThrProAlaPheSerPropheValArgValGluL 1602
12761 CCACGCGCATCGCGCGAGCGCGCATGATCT..... 12790
1602 ysArgAspAlaPheThrIleCysThrValValAsnSerProGlyAsp 1618
12791GCCCCATCCCCCGCGCTGTCT 12814
1619 AlaProLysProHisArgLysProSer..... 1628
12845 TCGCCAGCGCGCGCGCTCGCGCATCTATCTACCGCGCTACTATGGA 12864
1629SerAlaSerSerSer..... 1633
12865 CTGCTCGATCTCGGCACTGAAACCAATCAACGTGTCTCTATCCATGC 12914
1634SerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1645
12915 GCGCGCGCGCGCTCGGAGCGCGCGCTTCTAGCTCGACGCGACCTCG 12964
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPr 1659
12965 GCGCGAGGTCTTTCGCCAGCGCGCTCGAGGAAAGTGGAGCTCTCGCG 13014
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProValV 1675
13015 CGCGCTCG.....GTTTCGAGCTCGGACCTCGCTCGCTCGCTCG 13052
1675 alSerLysAla...LeuSerThrSerCys 1683
13053 TGACCTGGCTTCGAGCAGCAGCTTCTCG 13081

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-729-214-6

seq_documentation_block:

; Sequence 6, Application US/08729214

; Patent No. 5817502

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Hill, Dwight Steven

; APPLICANT: Ryals, John Andrew

; APPLICANT: Hammer, Phillip E.

; APPLICANT: van Pee, Karl-Heinz

; APPLICANT: Kirner, Sabine

; TITLE OF INVENTION: Genes for the synthesis of

; antipathogenic substances

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ciba-Geigy Corporation

; STREET: 520 White Plains Road

; CITY: Tarrytown

; STATE: NY

; COUNTRY: USA

470 rArgThrProGluGlnHisLysSerGlnHisCysSerProGluGlySerg 487
:: : ||||| :: | | | | |
8590 CTGCAGCCCC.....ACGACGCCCTGTCC.....TCGA 8618

487 lYTy-SerAlaGluProAglLyThrProLeuSerGluProPro..... 501
:: : :::::||::| ::||| ::||| ::||| ::|||
8619 ATCTGCTTGGCAGGCCCTCGAAGACGCCGGCATGTCCCCCACCTCA 8668

502SerSerThrProGlnSerThrHl 509
:::||||| ::||| ::||| ::|||
8669 AGGATTCOCCACCGCGCTCTCTCGGCATCGCGGCACGAATACGCA 8718

509 sAlaGlu.....ProGlnGluAlaA 516
|:||||| ::||| ::||| ::||| ::|||
8719 TTGCGAGGAGCAGACACGAGAATCGAGCTTAGCCCTTCACAGGCA. 8767

516 spYrLeuSerGlySergLiuAspProLeuGluArGsrPheLeuTrCys 532
||||| ::||| ::||| ::|||
8768CGCGGGGTCTTTGCCGGGGCGCTTGCCCTACA 8803

533 AsnGlnAlaArgLysSerProAlaArgValAsnSerIasnSerLysAlaLyl 549
::::||| ::||| ::||| ::||| ::|||
8804 CGTCGGCTCGAAGGGCCG...CGCTCTCGGTGCACCGCGCTGCTCC 8850

549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSart 566
|||||::||| ::||| ::||| ::||| ::|||
8851 TCCTGCTCTGTCCCTCCACTCGCTGCCAGACCCTTCGCACAGGGCA 8900

566 hrlYSexAspAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
::: ||| ::||| ::||| ::|||
8901 GTGCAACCTCGCCTCGCGGGCGTCTCGTCTATGGCTCCCGCGAGG 8950

583 SerPheSerLysPheValala..... 589
::: |||||::||| |||
8951 GCTGTGCTCTCTTCTCCGCTCGCTGTGGCCGCGACGGCGCTCC 9000

590GlyGluArgAspCysP 595
:::|||||
9001 ANGACCTTCTGGCACAACGCCGACGGCTAGCGACCGGAGAGGCTCAT 9050

595 roArgLeuLeuLeuSerAlaLeu.....AlaGlnGluAspLeula 608
::||| ::||| ::||| ::|||
9051 CGTCTTGCCCTCGAGCGCTCGGTGAGCGCTTCGCCCGCGACGCCGCTCC 9100

609 serGluIleLeuglyLeuGlnGluAlaIleGlyGlyLysAlaAspLysAl 625
|||||:: ::||| ::||| ::|||
9101 TCTTCGCTGTCTCGCGCACCGGCATCAACACGACGCGCGGTGACG 9150

625 a.....TrpAlaGluAlaProS 631
|.....
9151 GGTAATCAGCGCCCACACGGCACCTCCACAGAAAGGTCTCGCGCGCG 9200

631 erLeuVallyAspSerSerLysProPhoPheSerLeu..... 643
::: ::||| ::||| ::|||
9201 GCTCCACGACGCGCGATCACCCCGCGAGCTTCGAGCTGTGTAGTGGCC 9250

644GluAsnHisSerAlaCysLeuAspSerVa 653
|||||:: ||| |||
9251 ATGGCACCGGCACCTCTCTTGGAGACCCCATGAGGTG..... 9289

653 lAlaLysSerAlaIrrProArgProGlyGluProGluAlaLeuProAsps 670
::: ||||| ::||| ::|||
9290AAGCCCTGGCG...CGCTCTACGCGACGACGACCG.... 9325

670 erLeuGlnLeuAspLysGlyAlaAsnAlaLysAspPheSerProGlyLeu 686
|||::||| ::|||
9326CTGAAAAGCTCTCTCTCTCGCGG 9349

687 PheLiuAspProSerValAlaPheAlaThrProAppProlystTh 703
::: ||||| ::||| ::|||
9350 CGTTCAGAACCAATCGGCCCATCTCGAGCGCGCTCTCGCGCTCGCGGCG 9399

703 rglyProLeuSerPheGlyThrIysProThrLeuGlyValProAlaProA 720
7490 GTCCGCAAGATCG.....TCGCCTCCTCCGATCAGTCAGCGCTCCGCC 9443
720 spProThrThrAlaAlaPheAspCysPhePro.....AspThr 732
9444 CACCTTCACACAGGCGCGCGGCAT.....CCTTGATGATGTCGATACA 9489
733-ThrAlaAlaSerAlaAspSerAlaAsnProPheAlaTrpProGluG 749
7490 CTGCGCATCGAGCTGCTGTATACC.....CGAGGTCCTGGGCCCGCCA 9533
749 uasnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeu 766
9534 CGAA.....GATACGATCCCGC..... 9552
766 hrIysGlyLeuGluGluGlyGlyLysAlaSerAspGlyIleSerLysGly 782
9553CGCGCGCGGCTTCGGCTTCGGACTTCCTCGGCAC 9588
783 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGluG 795
9589 AAGCCCGACGTCATCCTCGAGGAGGCTCCCGCGCGCTG.....TC 9629
795 uGluAspProGlyGluLysValaSerLeuProGlyAspPheLysG 812
9630 GGGCGAGCGCGCCACTCACAGAGCGCTCGCGACCGCTCCCGCGCGCT 9679
812 lngluGluValGlyValLysGluGluAlaGlyGlyLeuLeuGlnCys 828
9680 GTGCGTGTCTGTTCGCGCAGAGGAGCGCGCTCGCGCCAGCG 9729
829 ProGlu.....ValAlaLysAlaAspArgTrpLeuGluAs 840
9730 AAGCGCTCCGCGACCACTCTCGCCACGACGACTCGCCCTATCGA 9779
840 pSerArgHisCysCysSerThr.....AlaAspPhe.GlyAspLeuPro 854
9780 TGTGGCTATTGCGAGGCGACCAACCGCGCCACTTCGAGCACCGCGCG 9829
855 LeuLeuProThrSerArgLysGluAspLeuGluAlaGluGluLut 871
9930 CTCTCTCGCGCGAGCGCGAGCG..... 9853
871 rSerSerLeuCysGluLeuLeuGlySerProGluGlnArgProGlyMetG 888
9854 AGCTCTCTCGCGCTGACTGCTGCTCGCCACGACGAGCGCG..... 9895
888 lnAspProLeuSerProLysAlaProLeuIleCysThrLysGluGluVal 904
9896CCCGAGCACCGCTTCCTGCGCGGAGCGGAGCGAC 9931
905 GluGluValLeuAspSer.....LysAla.....GlyTrp 914
9932 GCAAGCTCTCTGCTGCTTCCTCGGCACAGGCTGCGAGTGGGAGGATG 9981
914 pGlySerProCysHisLeuSerGlyGluSerValIleLeuLeuGlyPro 931
9982 CGCCTCTGCTGCTGACTCTCGCGGCTTCCTCGCGCTCAGCTCGAGC 10031
931 hrValGlyThrGluSerLysValGlnSerTrpPheGluSerSerLeuSer 947
10032 ATGCGAGCGCGGCTGCTCTCTCAGCTG..... 10060
948 HisMetLysProGlyGluGluGlyProAsp.....GlyGluArgAlaPr 962
10061AGTGGAGCTGCTGCGGCTCTCGCGCGACGAGGCGCC 10101
962 oGlyAspSerThrThrSerIspAla..... 970
10102 CCTCTCTGACCGCGTCTGAGCTGCTACACCGCGGCTCTTTCGCGTAT 10151
971SerLeuAlaGlnLysProAsnLysProAlaVal 981

10152 GGTCTCCCTGCGCCGCTCTGCGCTGCGTGAAGCCGCGCCG 10201
982 ProgluAlaProIleAlaLysGluProArgLysSerLe 998
10202 TCGTCGCGCAGCCAGGAGGAGATGCGCGCTCTGCGAGCCCT 10251
998 uArgSerArgValHisArgGlyLeuProgluAlaGluAspSerProC 1015
10252 CTCTCCCTGAGAGAGCGCGCGCATCGCG.....CCCT 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProIu 1028
10287 GCGCAGGAAAGGCTCACACCGTCGCGGACAGCGCGCATGCGCGC 10336
1029 SerCysThrGlyProProGlnGlyLeu.....MetGluVala.. 1041
10337 TCGAGCTCGCGCGCTCGACCTCCAGACCTACCTGCTCCGCGGCGC 10386
1042GlyAlaProGlyArgGlyAlaSerGluG 1051
10387 AGGCTCTCCACCGCGCGCGCTCACACCGCGCTACCTGCTATCGCG 10436
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGluProArg 1067
10437 CGAGCCCGCGCGCTCGACGCGCTGCTCACGCTCTCACCGCCACAGG 10486
1068 ...ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080
10487 TGTTCGCGCGCAAGATCCGCGTGAATACGCTCCACTCCGCGCCAGATG 10336
1081ProAspLysLeuGlyG 1086
10537 GAGCGCTCCAGAGACGCTCGCGAGGCTCAGCAACATGCTCTCGCG 10586
1086 LysGluAlaArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
10587 GAGCTCGAGCTCCCTCTTATT.....GACCGCTCA 10618
1103 ProLysAlaLaser...SerProSerAspProAla..... 1113
10619 CCGGACACAGGCTCGACGCTCGACGCTCGACGCGGCTACTGATGCA 10668
1114AlaLeuProValAlaSerAspSerProMetGlySerLysT 1128
10669 AACCTCGGCAACCGCTCTGCTCTCGACGCGACGACGCGCTCTCGA 10718
1128 hrLysGluThrAspSerProSer..... 1135
10719 CGATGGGATCGCTCTCGCTCGAGGTCAGCCCCATCCGCTCTCGACG 10768
1136 ThrProGlyLysAspGlnArgSerMetIleLeuArgSerArgThrLysT 1152
10769 TCGGCTCGCGGAGACGCTCGAGAGGCTCAACGCTCGATCCCGCTCGATC 10818
1132 hrGlnGluIlePheHisSerLysArgArgProSerGluGlyArgLeu 1168
10819 GG.....CTCCATTCGACGAGAGAGGCACTCG..... 10849
1169 ProAsnCys.....ArgAl 1173
10850 CCGGCT 10899
1173 aThrLysLysLeuLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190
10900 GACTGAGAGACTCTCTCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 10949
1190 erSerSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203
10950 CACCTACCCCTTCACGAGAGAGGCTTCGCTGAGCTCTCACGCGAGC 10999
1204 ArgVal..... 1205
11000 AACGCTTCGAGAGCTGCTCTCCGCAAGCTTGACCTGCGCCGACATCCG 11049
1206ProLysProGlyAla..... 1210
11050 CAGCTCGGCGCGCGCGCTGCGCTTCGCGAGACGCTGCTCTCTCTTAC 11099
1211GlySerLysLeuSerAspArg 1217
11100 AAGGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGGCGACCCACCTCTACCGGCGACGCGCTTCTCGAGCTCGCCCTGAC 11199
1229MetAlaProValProThrLys.....LysArg 1238
11200 GTCGCGCACCGCGCTGCGCTGACACGCTGGAAGGCTCACGCGAGAGC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaSerAlaMet 1254
11250 CCTCTGCT 11299
1255 Gly..... 1255
11300 GCGCGCTGACGACGAGAGAGAGAGGCGCTCTCTCTCTCTCTCTCTCA 11349
1256GluMet 1258
11350 GAGGAGCGGCTTCAGATGGCCCGGAGCTCGGACGCGACGCGGCTCTG 11399
1258 LysArgArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274
11400 CTGCGCGCGGACCCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11443
1275 LeuProArg.....LysProse 1280
11444 CTCTCTCTGAGTGCATCCCGGTGAGCTCGAAGGCTCTACGACCTCTC 11493
1280 rPro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCGAAGCTCGGCGCTGCTACGCGCGGAGATTCAGAGGCTCTCTCTCT 11543
1290 InSerSer.....His 1293
11544 CTACAGAGGCGGCGAGAGCTCTTTCGGAAGCAAGCTCCCGGAGACG 11593
1294 ProArgArg...ProProPheLeuProGlnAlaArgLeuSerAlaAlaPhe 1309
11594 CCGAAGAGATGCGCGCGGCTTGGCC...TCCACCTCTGCGCTCTGAC 11640
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgGlyLysGlyLys 1326
11641 AAGGCGCTTCGATGACCTGCTTGAAGAGAGACAGAGAGGAGGAGCTCG 11690
1326 eu..... 1326
11691 TCGGCGCT 11740
1327LysLeuGlnAlaIleValGlnLysIleThrSerProSer..... 1339
11741 CCTGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11790
1340LeuLysLysPhe 1344
11791 GTCGCGGAGCGCGCGAGTACGCTCTGCTCTCTCTCTCTCTCTCTCTCT 11840
1344 LacyS..... 1345
11841 CATGCGGAGAGCTCGCGCGGAGCTCGGACCGCGGAGCTCTCTCTCTCA 11890
1346LysAlaPro..... 1348
11891 ATGATGCGCT 11940

1348 1348
11941 CGCCTGCGCCGCGGCGGCGCTCTTCGCGACAGCGGCCACGATCT 11990
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
11991 CGCGCTGACGCCCGCGCTGCGCGTACGCGGCGCTGCTGCTCGGAA 12040
1361 SerAspHisArgGlyLeuLys...GlyAlaGlySerProValGI 1376
12041 GCGCGCTGACGCGGCGCTGCTGCTGCGCGCTGCTGCTGCGCC... 12085
1376 yValGluGluLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392
12086CCTTCATCATGACGCGGCGGCGCTGCTGCGGCGCGCGCA 12131
1393SerGlyAlaAspProLeuCysArg.....AsnProThrAsn 1404
12132 GCGCACCGCGCTGCGCTGCGCTGCTGCGGCGCTGCTGCGGCGCGAC 12181
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerThrAs 1421
12182 GCGTGCCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12231
1421 pCysPheLysThrGluAlaPheThrSerProGluAlaLeuGlnProGly 1438
12232 CACACGAGAGGAGGCTGACGAGGAGCTG.....CTCAGCGCGCGCTG 12275
1438 LysThrAlaLeuAlaProLysLysArgSerArgLysGlyAlaGlyAla 1454
12276 GGGGCTGCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12325
1455 HisGlyLeuSerLysGlyProLysGlyLysArgProTyrLeuGlyProAl 1471
12326 TGTGAGCATG.....ACCTCAGGAGGCGCTGCGGCGGCGGCGG 12366
1471 aLeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
12367 CTAGGCGCGCTGCGACAGAGAGAGCGGCTGCGGCGGCGGCGG 12416
1488 LuAspAsnSerGlyGly.....GlyGlyLys.....LysProLysMetGluGI 1493
12417 ACCCTCATCCGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCT 12466
1494GlyGlyLys.....LysProLysMetGluGI 1502
12467 CGCGCGACGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12516
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518
12517 GAGCGCTGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12560
1518 LysThrArgAlaGlnLysGlnProGlyHisThrAsnTyrSerTyrSer 1534
12561 CGCACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12610
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
12611 GCGATGCTGCGACACCGCTGCGGCGGCGGCGGCGGCGGCGGCG 12658
1544 aLysAsnThrThrSerProCysLysGlyAlaLysArgArgArg 1561
12659TCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12692
1561 LysGlnGlnValLeuProLeuAspPro..... 1569
12693 CGGTCCAGGTGCTCCGATACACCGTAGGCGGCGGCGGCGGCGG 12742
1570 ...AlaGluProGluLeuArgLeuLysLysTyrLysSerSerCysLysArgLe 1585
12743 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12760

1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602
12761 .CGACGCGCATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12790
1602 ysArgAspAlaPheThrThrIleCysThrValValAsnSerProGlyAsp 1618
12791GCCCATCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12814
1619 AlaProLysProHisArgLysProSer..... 1628
12815 TCGCCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 12864
1629SerAlaSerSerSer..... 1633
12855 CTGCTGATCTCGGCGCATCTGAACCCATCATCATCTCATCATCT 12914
1634SerSerSerSerSerSerSerSerSerSerSerSerSer 1645
12915 GCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 12964
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPr 1659
12965 GCGCGGAGGTCTTTCGCCAGCGGCGGCGGCGGCGGCGGCGGCG 13014
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675
13015 CGCGCTG.....GCTTCGAGGATGCGGCGGCGGCGGCGGCGG 13052
1675 alserLysAla...LeuSerThrSerCys 1683
13053 TGACCTGGGCTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 13081

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-028-934-6

seq_documentation_block:
; Sequence 6, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Phillip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyrimidin Biosynthesis Genes and Uses
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwalis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241

REFERENCE/DOCKET NUMBER: CGC1506/C1P7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO. 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 28958 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOHETICAL: NO
AMTI-SENSE: NO
US-09-028-934-6

alignment_scores:
Quality: 493.00 Length: 2095
Ratio: 0.593 Gaps: 102
Percent Similarity: 39.666 Percent Identity: 22.291

alignment_block:

US-09-508-821b-6 x US-09-028-934-6

Align seg 1/1 to: US-09-028-934-6 from: 1 to: 28958

62 GLYAlaGlyThrProSerGlyThrAlaAlaValAla.....Al 75
|||||
7609 GCGCGGACAGCAAGAGGCGCTAGCGCGCCCAACCGCTTCCTGACGC 7658
75 AAspLysTyrHisArgGlySerLysAlaLeuProThrGlnGln..... 89
|
7659 CCTGCGGAGCATCGCGCGACGCGCTGAGTACACAGCAGCTCGGCGCT 7708
90 ..GlyLeuGlnGly ArgProAlaPheProGlyTyrGlyValGlnAspSe 105
|||
7709 GGGCGCGCTGCGGCGCGCGCGCATGCGCACCGACGCGGAGCCGAC 7758
105 rSer.....ProTyrProGlyArgTyrAlaGlyGln 116
|||
7759 CTCGACAGCGCGCTTCGCGGATGCGCGCTTCGCGCGCTTCGCGCGC 7808
116 LuSerLeuGlnAlaTyrPheGlyAlaProGlnProProGlnPro..... 131
|||
7809 GCTGCGCTGCGCTTCGAGACGACGACGACGACGACGACGACGACG 7858
132GlnProLeu.....ProAlaGln 137
|||
7859 TCGACTGCGCGCGCTTCGCGCTTCGCGCGCTTCGCGCGCGCGCGC.. 7906
137 YValAlaLysTyrAspGluAsnLeuMetLysTyrAlaValProProS 154
|||
7907CGCTCCTGCGCGCATTCGCCGGA 7928
154 eArgGlnTyrAlaGlnGlnGlyAlaGlnValProPheArgThrHisSer 170
|||
7929 GCGCGACGCGCGCTTCGAGACGACGACGACGACGACGACGACG 7960
171 LeuHisValGlnGlnProProProGlnGlnProLeuAlaTyrProLy 187
|||
7961CGTCTCCGAGCAGGCGCGCGC..... 7984
187 sLeuGlnArgGlnLysLeuGlnAsnAspAlaSerProLeuProPheP 204
7984 7984
204 roGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSerSer 220
|||
7985CGGACCTCCGCG 7996
221 ThrTyrSerSerSerValGlnGlnGlyGlyGlnGlnAlaHisSerTyrLy 237
|||
7997 ACAGCTCCGAGCGCGCTTCGAGACGACGACGACGACGACGACGAC 8046

237 sSerCys.....ThrAlaProThrAlaGlnProH 247
|||
8047 CTGTCGCGCGACAGAGAGCGCCCTTCGCGCGCGAGAGAGCGCCCTCA 8096
247 IsAspArgProLeuThrAlaSerSerSerLeuAlaProGlyValArgVal 263
|||
8097 TGTCCAGCCCGGACAGAGGCTTCCTCGATC..... 8125
264 GlnAsnLeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnI 280
8125 8125
280 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeuGlnSerArgH 297
8125 8125
297 IsHisAlaGlnGlnThrLeuHisTyrGlnAsnLeuAlaLysTyrGlnHis 313
|||
8126TGGCTCTGATTCGCTCA 8143
314 TyrGlyGlnGlnGlnGlnGlnGlnGln.....ProAspAlaAlaVal 328
|||
8144 TGGCGCTGAGAGCTTCGCGCGCGCTTCGAGACAGCGACGCGCATCAAGCTC 8193
328 lArgThrProGlnGln.....TyrTyrGlnThrPheSerP 340
|||
8194 CCGGCGACCGCTTCGAGACGATCCCTTCCTGATGAGTGGCGCTCTT 8243
340 roSerSerSerHisSerProAlaArgSerValGlyArgSerProSerTyr 356
|||
8244 CTTCGCGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8290
357 SerSerThrProSerProLeuMetProAsnLeuGlnAsnPhe...ProTy 372
|||
8291AGCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8328
372 rSerGlnGlnProLeuSerThrGlyAlaPheProAlaGlyLeuThrAspH 389
|||
8329 AGCGACAGCGCGATGCGCGCTTCGAGAGCGCGCTTCGCGCGCGCGCG 8378
389 lSerHisPheMetProLeuLeuAsnProSer...ProThrAspAlaThr 404
|||
8379 CGTGGCGGATTCGAGCGCTTCGAGAGTCTCTGCGCGAGGAGCGCGAG 8428
405 SerSerValAspThrGlnAlaGlyAsnGlySerSerProGlnGlnLys...As 420
|||
8429 GCGTGC.....AGCCATTCGACAGAGCGCGCA 8454
420 pLysLeuProGlnAsnLeuLeuSerAspLeuSerLeuGlnSerLeuThr 437
|||
8455 TGGGATGCGCGCTTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 8504
437 lAlaThrLeuGlnValGlnAlaGlnLysSerAsnThrValGlnGlnLeu 453
|||
8505 CTAGCGCGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 8554
454 LeuSerTyrAlaAlaValProGlnLysLysGlyValLysAsnLeuValSe 470
|||
8555 CTTCTCTTGGCATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 470
470 rArgThrProGlnGlnHisLysSerGlnHisCysSerProGlnGlySerG 487
|||
8590 CTGACCGCGC.....AGCACCGCGCTGCTCC.....TGA 8618
487 lTyrSerAlaGlnProAlaGlyThrProLeuSerGlnProPro..... 501
|||
8619 ATCTGCGTGGCAGCGCTTCGAGACGCGCGCGCGCGCGCGCGCGCGCTCA 8668
502SerSerThrProGlnSerThrH 509
|||
8669 AGGATTCGCGCGCGCGCTTCGAGCGCGCGCGCGCGCGCGCGCGCGAG 8718

509 sAlaGlu.....ProGlnGluAla 516
|||||
8719 TTGGAGAGGAGGAGCAGCAGAGATTCGACCTTATGCTCCCAAGCA. 8767
516 spYrLeuSerGlySerGluAspProLeuGluArgSerPheLeuTyrCys 532
|||||
8768CCGCGGGTCTTTCGCGGGGCGCTGGGCTACA 8803
533 AsnGlnAlaArgGlySerProAlaArgValAsnSerAsnSerLysAlaLy 549
:|||||
8804 CGTCTGGCTGCAAGGCGCG...CGCTCTGGTGACACCGCTGTGCTC 8850
549 sProGluSerValSerThrCysSerValThrSerProAspAspMetSerT 566
|||||
8851 TCTCTCTCTGCGCTCCACCTTCGCTGCGCTGCGCAAGCTCCGACAGGGCGA 8900
566 hrLysSerAspSerPheGlnSerLeuHisGlySerLeuProLeuAsp 582
:|||||
8901 GTGCAACCTCGCTCGCGGGGCTGCTCGTCTATGGCTCCCGCAGG 8950
583 SerPheSerLysPheValAla..... 589
8951 GCTTCGCTCTCTTCCTTCGCGCTGCGGGCTTGGCGCGCAGCGCGCTCC 9000
590GlyGluArgAspCysP 595
9001 AAGACCTTCTCGGCAACGCGGACGCTACGACGCGGAGAGGCTCAT 9050
595 roArgLeuLeuSerAlaLeu.....AlaGlnGluAspLeuAla 608
|||
9051 CGTCTTGTCTCGAGGGCTCGGTGAGCGCTCGCGCGGAGGACACCGCG 9100
609 SerGluLeuLeuGlnGlnAlaIleGlyGluLysAlaAspLysAl 625
|||||
9101 TCTTCGCTCTCGCGGGCAGCGCATCAACGACGCGCGCTGAGC 9150
625 a.....TrpAlaGluAlaPro 631
9151 GGTATCAGCGCCCAACGCGACCTCCAGCAGAGGTCTCTCGCGCGC 9200
631 erLeuValLysAspSerSerLysProProPheSerLeu..... 643
9201 GCTCCACGAGCGCGCATCACCGCGCGAGCTCGACGCTCGTCTGAGTCC 9250
644GluAsnHisSerAlaCysLeuAspSerVa 653
9251 ATGGCACCAGCTCTCTGGGAGACCCATCGAGTGC..... 9289
653 lAlaLysSerAlaTrpProArgProGlyGluProGluAlaLeuProAspS 670
:|||||
9290AAGCCTGCGCG...CGCTCTACGCGGACGCGGACCG... 9325
670 erLeuGlnLeuAspLysGlyGlyAsnAlaLysAspPheSerProGlyLeu 686
9326CTGAAGACCTCTCTCTCTCGCGC 9349
687 PheGluAspProSerValAlaPheAlaThrProAspProLysLysThrTh 703
:|||||
9350 CGCTCAAGACCAACATCGGCATCTCGAGCGCGCTCTCGCTCGCGGC 9399
703 xGlyProLeuSerPheGlyThrLysProThrLeuGlyValProAlaProA 720
:|||||
9400 GTGCCCAAGATG.....TCGCTCTCTCGCGCATGACGCTCTCGCGC 9443
720 spProThrThrAlaAlaPheAspCysPhePro.....AspThr 732
:|||||
9444 CACCTTCACAGCGCGCGCGCAT.....CCCTTGATTGATGGATACA 9489
733 ThrAlaAlaSerAlaAspSerAlaAsnProPheAlaTrpProGlnG 749
|||
9490 CTCGCCATCGAGCTGTGTATACC.....CCGAGGTCTTGGCGCGCGCA 9533
749 uAsnLeuGlyAspAlaCysProArgTrpGlyLeuHisProGlyGluLeuT 766

9534 CGAA.....GATACAGTCCCGC..... 9552
766 hrLysGlyLeuGlnGlyGlyLysAlaSerAspGlyLysSerLysGly 782
:|||||
9553CGCGCGGGCTCTCGCTTCGGACTCTCCGGCACC 9588
788 AspThrHis.....GluAlaSerAlaCysLeuGlyPheGlnG 795
:|||||
9589 AAGCCCACTCATCTCTGAGGAGGCTCCGCGCGCTG.....TC 9629
795 uGluAspProProGlyGluLysValAlaSerLeuProGlyAspPheLysG 812
:|||||
9630 GGGGAGCGCGCCACCTCACAGAGCGCTCGGACCGCTCCCGCGCGCT 9679
812 lngGluValGlyGlyValLysGlnGluAlaGlyGlyLeuLeuGlnCys 828
:|||||
9680 GTGCGCTCTCTCTGCGCGAGGAGGCGCGCTCCGCGCGCGCGCG 9729
829 ProGlu.....ValAlaLysAlaAspArgTrpLeuGluAs 840
:|||||
9730 AAGCGCTCCGCGACCACTCTCTCGCGCGACGACGCTCGCTTATCGA 9779
840 pSerArgHisCysCysSerThr.....AlaAspPheGlyAspLeuPro 854
:|||||
9780 TGTGGCTTATTCGAGCGCACACCGCGCGCTCTCGAGCACCGCGCG 9829
855 LeuLeuProProThrSerArgLysGluAspLeuAlaGlnGluTyr 871
|||
9830 CTCTCTGCGCGCGCGCGAG..... 9853
871 rSerSerLeuCysGlnLeuLeuGlySerProGlnArgProGlyMetG 888
:|||||
9854 AGCTCTCTCGCGCTCGACTCGCTCGCGCGCGCGCGCGCGCGCG 9895
888 lnAspProLeuSerProLysAlaProLeuLysCysThrLysGlnVal 904
:|||||
9896CCGCGAGCAGCTCTCTCGCGCGGAGGAGCGCGCG 9931
905 GluGluValLeuAspSer.....LysAla.....GlyTr 914
:|||||
9932 GCAGGTCTCTCTCTCTCTGCGGCAAGCTCGAGTGGGAGGAGGATG 9981
914 pGlySerProCysHisLeuSerGlyGluSerValLeuLeuLeuGlyPro 931
:|||||
9982 GCGCT 10031
931 hrValGlyThrGluSerLysValGlnSerTrpPheGluSerSerLeuSer 947
:|||||
10032 ATCGAGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10060
948 HisMetLysProGlyGluGlyProAsp.....GlyGluArgAlaPr 962
:|||||
10061AGTGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10101
962 oGlyAspSerThrThrSerAspAla..... 970
:|||||
10102 CCT 10151
971SerLeuAlaGlnLysProAsnLysProAlaVal 981
10152 GGTCT 10201
982 ProGluAlaProLysGlnLysGlnProValProArgLysSerLe 998
:|||||
10202 TCGTCT 10251
998 uArgSerArgValHisArgGlyLeuProGluAlaGluAspSerProC 1015
:|||||
10252 CT 10286
1015 ys.....ArgAlaProValLeuProLysAspLeuLeuProGlu 1028
:|||||

10287 GCGAGAAAGGCTCACACGCTGCGGCGAAGCGGCGATGCGCGG 10336
1029 SerCysThrGlyProProGlnGlyLeu.....MetGlnGlyAla.. 1041
10337 TCGAGCTGCGGCGCTCGACCTCGACACCTACCTGCTGCTGCGGCGAC 10386
1042GlyAlaProGlyArgGlyAlaSerGlnG 1051
10387 AGGCTCTCACGCGCGCGCTCAACAGCCCGAGGCTACCTGCTATCCGG 10436
1051 LysLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlnProArg 1067
10437 CGAGCCCGCGCGCGCTCGACGCTCGACGCTCGACGACCAAGG 10486
1068ThrProGlyProProGlyLeuThrThrProAlaPro..... 1080
10487 TGTTCGCGCGCAAGTCCGCTGACGCTCGACCTCGACCTCGACGATG 10536
1081ProAspLysLeuGlyG 1086
10537 GAGCGCTCCAGAGAGAGAGCTCGCGGAGGCTACGCAACATGCTCTCG 10586
1086 LysGlnAlaArgAlaAlaPheLysSerGlyLysArgValGlyLysProSer 1102
10587 GAGCTCGAGCTCCCTCTTAT.....CGACCGTCA 10618
1103 ProLysAlaAlaSer.....SerProSerAlaProAla..... 1113
10619 CGGCGACAGAGCTCGACGCTCGACGCTCGACGCGGAGTACGATGGA 10668
1114AlaLeuProValAlaSerAspSerSerProMetGlySerLys 1128
10669 AACCTCGGCGAAGCCGCTCTCTCGACGCGAGCGGCTCTCTCGA 10718
1128 hLysGlnThrAspSerProSer..... 1135
10719 CGATGGCAATCGCTCTCTCGGAGAGTCAAGCCCATCCGCTCTCAAGC 10768
1136 ThrProGlyLysAspGlnArgSerMetLysLeuArgSerArgThrLys 1152
10769 TCGGCTCCGCGAGAGCTCGGAGCGCTCAAGCGCTCGATCCGCTCTGTC 10818
1152 hGlnGlnIlePheHisSerLysArgArgArgProSerGlnGlyArgLeu 1168
10819 GG.....CTCATTCGAGAGAGAGGACCATCTG..... 10849
1169 ProAsnCys.....ArgAl 1173
10850 CCGCT 10899
1173 aThrLysLysLeuAspAsnSerHisLeuProAlaThrPheLysValS 1190
10900 GACTGAGAGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10949
1190 erSerSerProGlnLysGlu.....GlyArgValSerGlnArgAla 1203
10950 CACCTACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10999
1204 ArgVal..... 1205
11000 AAGGCTTCGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11049
1206ProLysProGlyAla..... 1210
11050 CTGCTCGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11099
1211GlySerLysLeuSerAspArg 1217
11100 AGGCGGCT 11149
1218 ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
11150 TCGCGACACCGCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11199
1229MetAlaProValProThrLys.....LysArg 1238
11200 GTGCGCCACCGCGCTGCGCTGCGACACCGCTGAGAGCTGAGAGC 11249
1238 snLeuValLeuArgHisGlySerSerSerSerSerAlaMet 1254
11250 CCGCT 11299
1255 Gly..... 1255
11300 GGCCCGTGGAGACGAGAGAGAGAGGCGCTCTCTCTCTCTCTCTCTCT 11349
1256GluMetG 1258
11350 GAGAGCGCT 11399
1258 LysArgArgArgGlyLeuArgValProProProSerSerArgLysLeu 1274
11400 CTGCGCGGCGAGCCCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11443
1275 LeuProArg.....LysProSe 1280
11444 CTCTCTGAGTGCATCCCGGTGAGCTCGAAGGCGCTCTCTCTCTCTCT 11493
1280 Pro.....ProArgAlaMetAlaSerLeuProG 1290
11494 GCGACCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11543
1290 InSerSer.....His 1293
11544 CTACAGCGCGCGAGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11593
1294 ProArgArg.....ProProPheLeuProGlnAlaArgLeuSerAlaPhe 1309
11594 CGAAGAGAGTCCCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11640
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGly 1326
11641 AGCGCTCTGATGACTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11690
1326 eu..... 1326
11691 TCTGCT 11740
1327LysLeuGlnAlaIleValGlnLysIleThrSerProSer..... 1339
11741 CTTGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11790
1340LeuLysLysPheA 1344
11791 GTCTGAGCGAGCGCGAGTGAACCTCTCTCTCTCTCTCTCTCTCTCT 11840
1344 Lacs..... 1345
11841 CATCGAGAGAGAGCTCCGCGCGAGCTCTCTCTCTCTCTCTCTCTCT 11890
1346LysAlaPro..... 1348
11891 ATGATGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11940
1348 1348
11941 CCGCT 11990
1349GlyAlaSerProGlyAsnProLeuSer.....ProSerLeu 1360
11991 CGGCT 12040
1361 SerAspLysAspArgGlyLeuLys.....GlyAlaGlyLysSerProValG 1376
12041 GCGCT 12085

Page 40

[illegible][illegible]

```

26922  CCGGACACAGGCTGACAGGCTCCAGGCTCCAGGCGGCGCTACTGGATACGA 26971
1114  ....AlaLeuProValAlaSerAspSerProMetCysLeuTyr 1128
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
26972  AACCTCCGGCAACCCGCTCTGTCTTCAGAGCGACGACGAGCGCTCCGCA 27021
1128  hlySLDhlyAspSerProser ..... 1135
27022  CGATGGGACATCGCTCTCTGATGAGAGTACGACCCCATCCGCTGACGC 27071
1136  ThrProGlyLysAspAlaArgSerMetIleLeuArgSerAlaThrLeuTyr 1152
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
27072  TCGCCTCCGGGACACCTGACAGGCGCTACGACGCTGATCCGCTGCTGTC 27121
1152  hrgSLDhlyLeuPheHisSerTyrAlaArgArgProSerLeuGlyAlaGlu 1168
27122  GG.....CTCCATTTCAGAGAAAGAGGCAACCTCG..... 27152
1169  ProAsnCys.....ArgAla 1173
27153  CCGCGCTCTCTCTCTCTCTGCGGCGAGGCTCTTCACCCGAGGCTCGGCTC 27202
1173  aThrLysIleuLeuAspAsnSerHisLeuProAlaThrPheLeuValS 1190
27203  GACGTGAGAGACTTCTTCGCGGCGCCAGGCTCCCGGAGAAGTGTCCCTCC 27252
1190  erserSerProGlnLysGln.....GlyArgValSerGlnArgAla 1203
27253  CACCTACCCCTTCAGAGAGAGAGGCGTGTGCTCCGACGCTCCACAGGAGC 27302
1204  ArgAla..... 1205
27303  AAGGCTTCGAGAGCTGCTCCGAGGCTGACCTCGGCGGACCAATCCG 27353
1206  .....ProLysProGlyAla..... 1210
27353  CTGCTGGGCGCGCGCGCTGAGCTTCGCGGACGCGGCGTCTCTCTTAC 27402
1211  .....GlySerLysLeuSerAspArg 1217
27403  AAGGAGGACTCTCCCTCCGACAGACACCCTGGGCTCGAAGGCACTAGCCGT 27452
1218  ProLeuHisAlaLeuLysArgLysSerAlaPhe..... 1228
27453  TCGGCACACCCATCTACCCGAGGACCGGCTTCTTCGAGCTCCGCTCGAC 27502
1229  .....MetAlaProValProThrLys.....LysArg 1238
27503  GTCCGCCACCGGCTGGCGCTCGACACCCGTCGAAAGATCAACGCTGAGAGC 27552
1238  snLeuValLeuArgHisGlySerSerSerSerAsnAlaSerAlaMet 1254
27553  CCGCTCGGCTCCCATCGACGACGACACGCTGCTCTCCAGATCTCCGTCG 27602
1255  GLy..... 1255
27603  GCGCCCTGAGCAGCAGAGACGAGAGGCGCTCTCTTCATAGCCGACAA 27652
1256  .....GluMet 1258
27653  GAGAGCCGCTTCAGAGATGACCCCTGACTGCAACGCCAGGCGTCTCT 27702
1258  LysArgArgGlyLeuArgValProProProSerSerArgGlyCysLeu 1274
27703  CTGCGCGGCGACCCATCCCTCTCCGCGCATCTCCAGAGTGGC..... 27744
1275  LeuProArg.....LysProSe 1280
27747  CTCCTCTGAGATGCCATCCCGGTGACCTGAGAGGCTCTACGACCTTC 27796
1280  rPro.....ProArgAlaMetAlaSerLeuProG 1290

```


27797 GCCAACCTCGGGCTTTCCTACGGCCCGAGTTCACAGGCTCCGCTCCGT 27846
1290 InSerSer.....His 1293
: : : : :
27847 CTACAGCGCGGAGAGCTCTTTCGGAGACCAAGCTCCCGGAGCGG 27896
1294 ProArgArg...ProProPheLeuProGlnAlaArgLeuSerAlaAlaph 1309
: : : : :
27897 CCGAAAGGATCGCGCTGTTGCC...TCCACCTCGCTCGTGCAC 27943
1309 eGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgGlyL 1326
: : : : :
27944 AGCCCTCGATGACCTGGCTTTGAGACAGACAGAGGACGCTGC 27993
1326 eu..... 1326
: : : : :
27994 TCTGCCCTTCTCGTGGAGCGAGTCTCGCTCGCTCGTGCACCA 28043
1327LysLeuGluAlaIleValGlnLysIlePheSerProSer..... 1339
: : : : :
28044 CTTGGCGGTGGCTTCCACCGTCCAGAGGTGATCTCCCTCGATC 28093
1340LeuLysLysPheA 1344
: : : : :
28094 GTCTGCGCGAGCGCGAGCTGACCTCTTCCCTCGGTSCAAGCGCTGC 28143
1344 laCys..... 1345
: : : : :
28144 CATCGGACACAGCTCGCGCGCAGCTCCGACCCGCGAGCTTCCAC 28193
1346LysAlaPro..... 1348
: : : : :
28194 ATGATGGCTCTTCGGGTGCTGAGCGAGCTCCAAAGCCCCACTCA 28243
1348 1348
: : : : :
28244 CCGCTGCGCGCGCGAGCGGCTCTTCGCGACAGCGCGCCAGATCT 28293
1349GlyAlaSerProGlyAsnProLeuSer...ProSerLeu 1360
: : : : :
28294 CGGCTCGAGCGCGCTCGCGCTGAGCGAGCTGCTGCTCCCGAA 28343
1361 SerAspLysAspArgLysLeuLys...GlyAlaGlyLysProValG 1376
: : : : :
28344 GCGGCTCGACCGAGCGCTTCGCTCCGCGCTCGCTCGCGCC... 28388
1376 yValGlnGluGlyLeuValAsnValGlyThrGlyGlnLysLeuProThr. 1392
: : : : :
28389 ...CCTCATGATCGACCGCGAGCGAGCTCTCCGAGCGCCACGA 28434
1393SerGlyAlaAspProLeuLysArg.....AsnProThrAsn 1404
: : : : :
28435 GGCACCGGCTCGACCTCGCTCTTGAAGCTGCTGCGCGAGAC 28484
1405 ArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeuSerSerThrAs 1421
: : : : :
28485 GCCTCGCTCGCGCGCTGCTCTGTCACCGAGCGCGCTCGCCACC 28534
1421 pCysPheLysThrGlnAlaPheThrSerProGluAlaLeuGlnProGly 1438
: : : : :
28535 CACACGAGACGAGCTCAGACCTCG.....CTCAGCGCGCTG 28578
1438 lyThrAlaLeuAlaProLysLysArgSerArgLysGlyArgAlaGlyAla 1454
: : : : :
28579 GGGCTCGCGCTCGCGGAAAGTAGACACCGACCTCCCGCTTC 28628
1455 HisGlyLeuSerLysGlyProLeuGluLysArgProTyLeuGlyProAl 1471
: : : : :
28629 TCGTCGACATCG.....ACCTCAGCGAGCTCCACGAGCGCTG 28669
1471 aLeuLeuLeuProArgAspArgAlaSerGlyThrGlnGlyAlaSerG 1488
: : : : :
28670 CTAGGCGCGCTCGACACAGAGACCCAGCTCCGCTCCGCAAGGAA 28719

1488 luAspAsnSerGlyGly..... 1493
: : : : :
28720 ACCCTCATCCGAGGTGGCGACACCGCTCGACGAGCGCTCATCC 28769
1494GlyGlyLys.....LysProLysMetGluG 1502
: : : : :
28770 CGCGCAAGCACCCAGCTGGCGCTCCATATTCGACCAAGGACCTTC 28819
1502 uLeu...GlyProAlaSerGlnProGluGlyArgProCysGlnProG 1518
: : : : :
28820 GAGCGCTCGCTCGTGGAGCGCCCG.....AGGCCAGGCGCCCT 28863
1518 lnThrArgAlaGlnLysGlnProGlyHisThrAsnTySerSerTySer 1534
: : : : :
28864 CGCACAGGCAAGTCCGCTCGCTGCGGAGCGGCGGCTCAACITC 28913
1535 LysArgLys.....ArgLeuThrArgGlyArgAl 1544
: : : : :
28914 GCGATGCTGCTCGACACCTTGGCATGTATCGGGCGACGCGCCGCG... 28961
1544 aLysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgG 1561
: : : : :
28962TCGGAGCGGAGCGCGCGCATGCTTACTGAGT 28995
1561 lnGlnGlnValLeuProLeuAspPro..... 1569
: : : : :
28996 CGTCCAGGTGCTCCGATACACGATAGCGACCGGCTGATGGGGTCT 29045
1570 ...AlaGluProGluIleArgLeuLysTyIleSerSerCysLysArgLe 1585
: : : : :
29046 TCGCGCGACCTTGGTC..... 29063
1585 uArgSerAspSerArgThrProAlaPheSerProPheValArgValGluL 1602
: : : : :
29064 .CCACGCGCATCGCGCGCGCGCATGATCT..... 29093
1602 ysArgAspAlaPheThrThrIleCysThrValValAsnSerProGlyAsp 1618
: : : : :
29094GCCCATCCCCACGCTGCTCT 29117
1619 AlaProLysProHisArgLysProSer..... 1628
: : : : :
29118 TCGCCCAAGCGCGAGCGTCCCGCATCATCTATCTACCGCTTACTATGA 29167
1629SerAlaSerSer..... 1633
: : : : :
29168 CTCGCTGATCTCGGSCATCTGAACCCAAATCAGCTGCTCTCATCCATGC 29217
1634SerSerSerSerPheSerLeuAspAlaAlaGly 1645
: : : : :
29218 GCGCGCGCGCGCTCGGAGCGCGCGCTTCACTCGACCGCACCTCG 29267
1646 AlaSerLeuAlaThrLeuPro.....GlyGlySerIleLeuGlnPr 1659
: : : : :
29268 GCGCGAGGTCTTTCGCGCGCGCTCCAGGGAAGTGGAGCGCTCTCGC 29317
1659 oArgProSerLeuProLeuSerSerThrMetHisLeuGly...ProVal 1675
: : : : :
29318 GCGCTCG.....CTTCGACGATCGCGACCTCGCTCTCAG 29355
1675 alSerLysAla...LeuSerThrSerCys 1683
: : : : :
29356 TGACCTGGGCTTCGAGCAGCACTTCTCTGC 29384

seq_name: /cqn2_6/ptodata/2/lna/6B_COMB.seq:US-09-144-085-3

seq_documentation_block:
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.

[illegible]

12181 GCACCCGCCCCAGCGCGCACAGCAGCCATCCGGACGCATCGACGT 12133
703 703
12131 ATGTGCTCACCGCGCGCGCGCAAGCGTGGCATCCGTCCAGGAAGCGC 12082
704 GlyProLeuSerPheGlyThrLysProThrLeuGlyVal. 716
12081 ACGGCTCGCGACCTGCGCTCACCCAGTACTGGCGGCACATCAGCGATTC 12032
716 716
12031 TTGGCCACCGAGCTGCGCGGTACCGCGCTCCACAGSGTAOCTGGCGC 11982
717 ProAlaProAspProThrThr.....AlaAlaph 726
11981 TGCCATAGCGGCATCTCCGCGGCCACTCCGCTACTCTTCACGATCCGG 11932
726 eaSPcysPheProAspThrThrAlaAlaSerAlaAspSerAlaAsp 743
11931 TCCATTCGCGGCTGTGGATCGTGGACAGCGCCAGACCGCGG..... 11886
743 roPheAlaTrrProGluGluAsnLeuLeuAspAlaCysProArGtrPrgly 759
11885 ..TGCGCGCGGCTGGCGCTCGAACCGCTGCCACTGCCACCTCCAGCAGCGCG 11838
760 LeuHisProGlyGluLeuThrLysGluGluGlyGlyLysAlaSe 776
11837 CTTGTCGCCGCTCAGCACGCTGGCTGGACCGTGAAGCCCGCGATC 11788
776 rAspGlyLeSerLysGlyAspThrHisGluAlaSerAlaCysLeuGlyP 793
11787 CCAAGCGCGGCTCCAGCCCATCGCGCAGCG..... 11754
793 heGluGluAspProProGlyGluLysValAlaSerLeuProGlyAsp 809
11753 CCGCTGCGCTCGGCTCCGACGCTCCACGACA 11718
810 PheLysGluGluValGlyValLysGluGluAlaGlyLysLeuLe 826
11717 TCATCGCTCGCGCGCTGCG..... 11697
826 uGlnCysProGluValAlaLysAlaAspArgTrpLeuGluAspSerArgH 843
11696 ACCTGCGATCGACGACCGC.....GGCGCA 11669
843 isCysSerThrAlaAspPheGlyAspLeuProLeuProThr 859
11668 CACTAGCTGCTGCTGCCGCAAGCTCA.....GCAGCGCGGCACGT 11625
860 SerArgLysGluAspLeuGluAlaGluGluThrSerLeuCysGl 876
11624 GCGCAGCGCTCA.....GC 11611
876 uLeuLeuGlySer..ProGluGlnArgProGlyMetGlnAspProLeuSer 892
11610 TCTCTATGAGTGCCTCCAGACGCGCGCGG.....CCTACTCC 11570
893 ProLysAlaProLeuLeuCysThrLysGlu..... 902
11569 CCAGCGCTCCCACTGGCGGTACAGCGCACTTCCAGCGCAACACCGCG 11520
903GluValGluGluValLeuAspSerLys.....AlaGlyTrpG 915
11519 GCTGCGTGTATCTGCGCTCCAGCAGCGCTCTGCTCGCTCGCGCT 11470
915 LysSerProCysHisLeuSerGlyGluSerValIleLeuLeuGlyProThr 931
11469 TCGCGCAACACCACTCTCT.....CAACCCAG 11441
932 ValGlyThrGluSerLysValGlnSer..... 940
11440 GTC.....GAGATGCGCTGTCAGCGCGCTCGCACACTCGTCGAAAGCGCG 11397

```

941 .....TTPheGluSerSerLeuSerHisMetLysProG 952
      |||
      |||
11396 CAGGAGAACCGGGGTACACCTGGTAAATCTCTCCCACTCCAGGCCG 11347
952 LysGluGluProAspGlyGluArgAlaProGlyAspSerThrThsSer 968
      |||
      |||
11346 TGGCTCCCTGCCCCCGGAACAGCC..... 11321
969 ASPAlaSerLeuAlaGluLysProAsnLysProAla.....ValProG1 983
      :|||:|
      :|||:|
11320 ...CGCAACCTTCCCCCGGGCGCTCCGCTGCGCGCACACACCTGCCG 11274
983 uAlaProLeuLeuLysLysGluProValProArgGlyLysSerLeuArg 1000
11273 CGGCTGCAACCTCCGGAAGGGCCCTCAAGCCCTCCACGCGCTCCGAGACG 11224
1000 eArnArgValHisArgGlyLeuProGluAlaGluAspSerPro.CysAr 1016
      :|||:|
      :|||:|
11223 CTGGCGCGCTGGACCAAGTATGGCGCCA...GGCAATAGTGGCTCGATGACG 11177
1016 gAlaProValLeu.ProLysAspLeuLeuLeuProLusSerCThsGly 1032
11176 CGCGCGCGGTCTCACACAG.....CGGCGCACCGCA 11145
1033 ProGluGluLysMetGluGluAlaGlyAlaProGlyArgGlyAla 1049
      |||||:|
      |||||:|
11144 CCGCCCGCGTGG...CTCCAGACCACTTCGCCAGCGCGCGCGCGCGCC 11098
1049 rGluGlyLeuProArgMetCysThrArgSerLeuThrAlaLeuSerGlu 1066
      |||||:|
      |||||:|
11097 CCGACGAGAGCTCTGCTCCCGCG.....ACAGACAGACGGGAG 11057
1066 rArnArgThrProGlyProGluLeuThrThrPro.....Ala 1079
      |||
      |||
11056 CGGAGAACCTTCGCGCCCGCACAGGCTCGCGACCGCGCGGCTCGCGCA 11007
1080 ProProAspLysLeu..... 1084
      |||||:|
      |||||:|
11006 CCGCTCTCAGCGCGCGCTCTTGAGAGAGACATGCGCTTCGCGCGCTG 10957
1085 .....GlyGlyLysGluThrAla. 1090
10956 ATGCCAACAGACAGACAGCCCGCAAGCGGGGAGCGCTCGGGCGCGCG 10907
1091 ..AlaPheLysSerGlyLysArgValGlyLysPro.....SerProLys 1104
      |||
      |||
10906 AGCGCAGCGCGTCAAGGCGTCAACCACTGCACCGGCAACATCTGCCAG 10857
1105 .....AlaAlaSerSerProSerAsnProAl 1113
10856 AGATCTGGGGGTGCGCGCGAGCTGTCAACGTCGCCGCGAGGGCTCTCA 10807
1113 aAlaLeuProValAlaSerAspSerSerProMetGlySerLysThrLysG 1130
      |||
      |||
10806 TGGC.....GAACCGCGCCAGATTTTCACACTCC 10775
1130 LuThrAspSerProSerThrProGlyLysAspGluArgSerMetLeuLeu 1146
10774 GGGGATGCGCGCGCCCAACTCCAGGTGACCGAGTGTCTTGACCGCTC 10725
1147 ArgSerArgThrLysThrGluGluLeuPheHisSerLysArgArgAlaGpr 1163
      :|||:|
      :|||:|
10724 CAAAGC..... 10719
1163 oSerGluLysArgLeuProAsnCysArgAla.....ThiLysLysL 1177
      :|||:|
      :|||:|
10718 .CGAGCGGCGAGTGCCTCTCCCTGCTTGCCCGGACACCGCGCGAGCG 10670
1177 euleuAspAsnSerHisLeuProAlaThr..... 1186
      :|||:|
      :|||:|
10669 TTGCACCTTGATCGGATCGCCAGAGAGTACCCTGACCTGAGCATCGAGCATCGA 10620

```

1187 ...PhylsValasSerSerProhInlysglyArgValSerGlnAr 1202
10619 CCAATCGAGTCTGGAGCGCTCCAGCCCGCG 10588
1202 galaaTgValProlysProGlyAlaGly 1211
10587 TTCCGAGAGCGCGGCGCCACACACTTCGCTGGGCGGTCCGTTCCGGCGG 10538
1212SerIysIenSerAspArgProIenHisAlaIen 1222
10537 CGTGAACCCCGCTGGCTCCGCGCGCTCGATTCAGCCGCGCCCGCACGA 10488
1223 LysArgLysSerAlaPheMetAlaProValProThrIysLysArgAsnLe 1239
10487 CGCGCAGACACCCGCAATCCCTCGCGCTGCTGCCTCCGAA 10449
1239 vValIenArgHisGlySerSerSerSerSerAsnAlaSerAlaMetIyG 1256
10448GAGGATATGAGCACACACAC 10430
1256 IuMetGlyArgArgArgGlyIleuArgValIProProSerSerArgGly 1272
10423 CCGAGAGCGCTCGCGCGCGCGTGAAGCGTGGCGCGCTCGAGAACGCT 10380
1273 CysIleuLeuProArgGlyLysProSerProProArgAlaMetAlaSerLeuPr 1289
10379 TGCACCGTCCGCGCGGAGACCGACGACCGTGGCGACACAGAC 10336
1289 ogInserSerHisProArgArgProProPheLeuP 1301
10335 ACACACCCCGCGCGGCTCGGAGACACTGCACGCGCGCGCCAGAGCCAC 10286
1301 roGlnAlaArgLeuSerAlaAlaPheGln 1310
10285 CTGCGATCGCGCGCTCGATACGCTTCGCACGACGATGACACGACGAC 10236
1311 gLyAlaMetLysThrLysValLeuProProArgArgGlyArgGlyLeu 1327
10235 CGAGCAGCTGCACCGCGCTATCGACCGAACCGGAGGCGCTTGCA 10192
1327 sLeuGlnAlaIleValGln 1333
10191 GCCGAGAGGATAGATCCACACAGCGCGCGTGGAGAGCTCAGCGCGGTCCC 10144
10141 GTCCAGCGCATTCGATTCGTTGCTCGCGCTTCGCGCGATTCGATCA 10092
1334LysIleThrSerProSerLeuLysIysPheAla CysLys 1346
1346 sAlaProGlyAlaSerProGlyAsnProLeuSerPro 1358
10091 CTGCGTGGCGCGCAAGCGCACAGAACACACCGGTGTCCAGACCCCTTCACT 10042
1358 1358
10041 GGTCTGGACGACACTCCGCGCTCCCTCAGCGCGCTCCACAGACTCTCCAGC 9992
1359SerLeuSerAspLysAspArgIleu 1367
9991 AGCAGCCGGTCTCGCGATTCATCGCTCCGCTCCGCGCGGCTTATCC 9942
1367 1367
9941 GAAGAACCCCGCGTCAACAGATCCACGCTCCACAGACGAGGAGACT 9892
1368LysGlyAlaGlyLysArgProValGly 1376
9931 TCACGACGCTCTTCCTTCCTTCGCTCGGGGTAGATTCGATCGCTCC 9842
1377 ValGlnGlnGlyLeuValAsnValGly 1385
9841 AGCTCAACAGGTGCGGAGATCGCGCTTAGGATTCGCGCGCTCCAC 9792
1386 ...ThrGlyLysIleuProThrSerGlyAlaAspProLeuGlyAsArg 1401

9791 GAGCATTGGCCAGAGCTCTCCAGATGTTCTCGCTCCGCGCGCATGGCA 9742
|||||
1401 snProThrAsnArgSerLeuLysGlyLysLeuMetAsnSerLysLysLeu 1417
:|||||
9741 GCCCAGCGCCACGATCGTATCGCGCGCTCCGCTCCGGTCCAGGCTC 9692
:|||||
1418 Ser.....SerThrAspCysPheLys 1424
:|||||
9691 AGCGCGTCTGCTCCGCGCGCTCCGCGCGGATCAGCTCCAG 9642
:|||||
1424 sThrGluAlaPheThrSerPro.GluAlaLeuGlnProGlyGlyThrAla 1440
:|||||
9641 CAGCCAGCGGCTTACCTCCGCTTGGCTGGGTAATCGAAGATCAACGTC 9592
:|||||
1441 LeuAlaProLysArgSerArgLysGlyAla..... 1452
:|||||
9591 TTGTACCGACACCCCGCTCCGCTCGACCGCGCTTCAGAGTCCAC 9542
:|||||
1453GlyAlaHisGlyLeuSerLysGlyProL 1462
:|||||
9541 GCCATCAGAGTCCAGCCCGAGTCCAGAAAGCTTCGTCGGGTCCAG 9492
:|||||
1462 euGluLysArgProTyrLeu...GlyProAlaLeuLeuLeuThrProArg 1477
:|||||
9491 CCGCTCGGATCTTACGCCAGCAGCAGCGCGCTCTCGCTCGCCACGA 9442
:|||||
1478 AspArgAla.....SerGlyThrGlnGlyAla.SerGluasp..... 1489
:|||||
9441 GCTCGCGCAGCGCTCGCTCCAGCCAGCAGCGCGCTCTCGCTCGCCACGA 9442
:|||||
1490AnSerGlyGlyGlyLysLysPro 1498
:|||||
9391 TCCCGCAGCGCTTCTCCCGCTCCGCGCGCTTCGCTCGGCTGTCG 9342
:|||||
1499 LysMetGluGlu.....LeuGlyProLysSerGlnProProGln 1511
:|||||
9341 GCCCTCAGCGATGCGCGCTTGTCTGATCCGTCAGCAGCGGCT... 9294
:|||||
1511 uGlyArgProCysGlnProGlnThrArgAlaGlnLysGlnProGlyHis 1528
:|||||
9293 ...CGCGCGCGCGCGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCG 9248
:|||||
1528 hAsnTyrSerSerTyrSerLysArgLysLeuThrArgGlyAla 1544
:|||||
9247 TC.....CGGATCCCGCCACCGAGTGGC 9225
:|||||
1545 LysAsnThrThrSerSerProCysLysGlyArgAlaLysArgArgGln 1561
:|||||
9224 CCGCAGCCCACTCGAGCGCGCAAGCGCTTGTCCGGCGAGATCGGCA 9175
:|||||
1561 nGlnGlnValLeuProLeuAspProAlaGlnProGluLysArgLeu.... 1576
:|||||
9174 CCAG.....CCCGCGCTCCGAGCTCGGCTCGGCC 9143
:|||||
1576 1576
:|||||
9142 TCGTGTCCATCCCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCG 9093
:|||||
1577 LysTyrLysSerSer.CysLysArgLeuArgSerAspSerArgThrProA 1593
:|||||
9092 CGCGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9043
:|||||
1593 laPhe....SerProPheValArgValGluLysArgAlaPheThrThr 1608
:|||||
9042 CGTTCGCGCGCGGTAC..... 9026
:|||||
1609 IleCysThrValValAsnSerProGlyAspAlaProLysProHisArgly 1625
:|||||
9025CCGCGCTCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 8997
:|||||
1625 sPro.....SerSerAlaSerSerSerSerSerSerS 1638
:|||||

8996 AGCGATGTGCGCATAAAGACAGCAAGCATCCAGCTCCGCTCGCGCGCA 8947
1638 er.....PheSerLeuAspAla..... 1643
:|||||
8946 GCTCTTGAGAGTGCATGCTCCGTTGACCTTCGCGGAGCTCCGCGCC 8897
:|||||
1644AlaGlyAlaSer.LeuAlaThrLeuProGlyGlySerLysLeu 1658
:|||||
8896 AGCATCTCGCGCGGAGCTCTCGACCGCGCAGCGCGGA.....CTAT 8853
:|||||
1658 lnProArgProSerLeuProLeuSerThrMet..... 1669
:|||||
8852 CCGCGCGCATGCGCGCGCGCTCAGCGCGCTCTCTCTCTCTCTCTCT 8803
1670HisLeuGlyProVa 1674
:|||||
8802 CCCTCAGCGCGCTTCGACCTCGCTCGCATCATCGCAGCGCGCA 8753
1674 lValSerLysAlaLeuSerThrSerCysLeuValCysCysLeuGlnA 1691
:|||||
8752 AGCGTCACGCGG.....C 8739
1691 snProAlaAsnPhelLysAspLeuGlyLysCysGlyProTyrTyrPro 1707
:|||||
8738 GCCCGCGC.....CCTCAGCTCTCT 8719
1708 GluHisCys.....LeuProLysLys..... 1714
:|||||
8719 CCGAGAGCTCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 8669
1715LysProLysLeuLysGlyLysValArgProG 1725
8668 ACCAGGTGCTCTCTCCCGCTCCGCGCGCGCGCGCGCGCGCGCGCG 8628
1725 LuGlyThr.....CysGluGlu 1730
8627 AGGCTCCGCGCGCTCCCGCTCCCGCTGATGAGCAGCTGCGCGCG 8578
1731 AlaSerLeuProLeuGluThrLeuLysGlyPro...GluCysAlaAl 1746
8577 GCTTCCACCGCGCTCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 8528
1746 a.....AlaAlaThrAlaGlyLysProPro 1754
8527 ACAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8497
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-335-409-1
seq_documentation_block:
: Sequence 1, Application US/09335409
: Patent No. 6121029
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/335,409
: CURRENT FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-335-409-1
alignment_scores: 468.00 Length: 2079
Quality:

Ratio: 0.565 Gaps: 113
Percent Similarity: 39.827 Percent Identity: 22.799

alignment block:

US-09-508-821B-6 x US-09-335-409-1/rev ..

Align seg 1/7 to reverse of: US-09-335-409-1 from: 1 to: 68750

```

26 ArgLeuGlnAsnTyrArgGlnProSerGlnAlaGlyLeuSerGlyAsp 42
|||||
28171 CGGCTACGCCGCCACGACGCCACACGCG.....CG 28140
42 GlnArgLeuAlaAlaLysAspTyrTyrAsnProGlnProTyrProSer 59
|||||
28139 TCAGCGGCTACTCCACCGTGA..... 28118
59 YrGlnGlyGlyAlaGly..ThrProSerGlyThrAlaAlaValAla 75
|||||
28117 ..GAGCGCGGGCTGGGTGAACGCGCTGCTGTCAGACACACGACT.... 28074
75 LaAspLysTyrHisArgGlySerLysAlaLeuProThrGlnGlnGlyLeu 91
|||||
28073 .....CGGCGCTCCCGGCTCCGCCACATCCACTCCGCGCA 28038
92 GlnLysArgProAlaPheProGlyTyr..... 100
|||||
28037 GCGGCGGCTCCAGCT...CCGCTCGAAGAGCGCCACGACCGGTCGAAAC 27991
101 .....GlyValGlnAsp 105
27990 GCCTCCGGAACCTGGCCACGCGCCGCAAGCCCGCGCCATGCCCG 27941
105 eSerProTyrProGlyArgTyrAlaGlyGlnGlnSerLeuAla... 120
|||||
27940 CATTGCGCGCCCTGTCCGCTGAAGAGAACGCACTTGCCTGCGCGACG 27891
121 .....TPrGlyAlaProGlnPro.....ProProGln 130
|||||
27890 AGCTCGGATGACAGCGCGCGCCCGCGCGGCTGCGCCCTGCGCGACG 27841
130 nProGlnProLeuProAlaGlyValAlaLysTyrAspGlnAsnLeuMet 147
|||||
27840 GCGGAAAGCGCGCCACGACG.....CCT 27815
147 yLysThrAlaValProPro.....SerArgGlnTyrAla 158
|||||
27814 GCGGAGCTCACGCCACGCGAGCGGCTGTCATCGGCTGCGCGCTCG 27765
159 GlnGlnGly.....AlaGlnValProPheArgTyrHis 169
|||||
27764 TCGCAGGCTGAAGCGACGCTCCCGAGCGCGAGCTCGGATGATGCTC 27715
169 sSerLeuHisValGlnGlnProPro.....ProProGlnGln 181
|||||
27714 AGGTGCTCGCGCAGCGCGCGCTGCGCATCGAGCGCGCTGCGCTT 27665
182 .....ProLeuAlaTyrProLysLeuGlnArgGlnLysLeuGln 194
|||||
27664 GCCCGACAGCACAAAGCTCCGCGAGCGCTCCGCGCGGAGGACACA 27615
195 AsnAspIleAlaSerProLeuProPheProGlnGlyThrHisPheProGln 211
|||||
27614 GCTCCACCGCGCGCGCTCTCCACACCA..... 27585
211 nHisSerGlnSerPheProThrSerSerTyrSerSerValGlnG 228
|||||
27584 .CATGCGGTGCTGCTGCTCATCCGACAGAGCTCAGCCCGCGAGAGC 27536
228 LysGlyGlnGlnGlyAlaHisSerTyrLysSerGlyThrAlaProThrAla 244
|||||
27535 CGGCGGCTCGCTGCGC.....GCCAGGACACGCTGCGC 27501
245 GlnProHisAsp.....Ar 249

```

```

27500 GAGCCACCGCCACCGGACGCGCTCCGACGAGACCGGATGAGCTCG 27451
|||||
249 gProLeuThrAlaSerSerSerLeuAlaProGly GlnArgValGlnAsn 265
|||||
27450 CCGAGCTCGCGCTGCGCTGATATTTCCTTGGCCCGGCGAGAGCGC 27401
266 LeuHisAlaTyrGlnSerGlyArgLeuSerTyrAspGlnGlnGlnGln 282
|||||
27400 CTGAGG.....CAGCGCGCCAGCGCGCGCGGCTCCATGAGGCCAA 27357
282 n.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 294
27356 GATTGCGCTTGGCGGCTCCAGGATCAGCGGCTGCTCCAGAGGCGCT 27307
294 LysArgHisHisAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 310
27306 TGGCGATGAGCTGCTGAGGAGG.....CACTCGATGGGTGCCCG 27260
311 Tyr..... 311
27259 CGCGCTCCGCGCTCCGCTGCGATTCAGAAATGAGTGGCGGAAACA 27210
312 .....GlnHisTyrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 319
27209 CCGCTCGCGCGCCACGCGCTGCGGAAACGCGCTTCCGCGAGGCGC 27160
319 LysGlyTyrGlnProAspAlaAlaVal.....ArgThr 330
|||||
27159 CTGGGACGTGTCAGCGCTGCTGCGGCGGCGGATGATGATGCGCTAC 27110
331 ProGln...GlnTyr..TyrGlnThrPheSerProSerSerHisSer 346
|||||
27109 CCGGATCAGCGCGAGATGAGGCTCGCGGCTGCGGCTGCGGAGAGCC 27060
346 roAlaArgSerValGlyArgSerProSerTyrSerSerThrProSerPro 362
|||||
27059 GCTTGAACACACACGCGCGCGCGCTGCGCGCTGCGGCTGCGCGC 27010
363 LeuMetProAsnLeuGlnAsnPhePro..... 371
27009 TCGCGGAGAGCTTGTGACCGCGCGCGCGGCAAGTCGCAATCGC 26960
372 .....TyrSerGlnGlnProLeuSerThrGlyAlaP 382
26959 TGAAGACGCGGAGGCTTGGCGAGAGAGCATGTGACCCG.... 26914
382 heProAlaGlyLeuThrAspHisSerHisPheMetProLeuLeuAsnPro 398
|||||
26913 ..CCACGAGGCTTGTGCGACTCGCCCGCTGCGAGCTGCGAGCGC 26866
399 .....SerProThrAspAlaThrSerSerValAspThr...GlnAl 411
26865 AGGTGACGCGCCACGAGGAGAGAGACGCGCTTATCCACACAGGCT 26816
411 agLysAsnGlyValLysProLeuGlnLysAspLysLeuProGlnLysLeu 428
|||||
26815 CGGCGGCTGCGAGCGCCAGGAAATGATGACCGCTCCGCGCGAGCGTGA 26766
428 eAspLeuSerLeuGlnSerLeuThrAlaLeuThrLeuGlnValGlnAsn 444
|||||
26765 GCAGTGTGCGCGTGGCG..... 26749
445 lIleSerAsnThrValGlnGlnLeuLeuLeuSerLysAlaAlaValProGln 461
|||||
26748 .....CTGTGAACCGCGCGCTCCATCGGC 26723
461 nLysGlyGlyValLysAsnLeuValSerArgThrProGlnGlnHisLys 478
|||||
26722 GAGTTCCTGACCGCGCTGCGATTCGTTGGGCGCGCGCGACACAC 26674
478 eArgHisGlySerProGlnGlySerGlyTyrSerAlaGlyProAla Gln 494
|||||

```

26673ACGCGCGTGGGGTCTCGCGCAGCGCGCGCGGCTCTCTGG 26635
494 yThrProLeuSerGluProSerSerThrProGlnSerThrHis...A 510
26634 CCGCGCGCTCGATCGCTCCACCTCACTCCACACGACGCGCTGTG 26585
510 laGluProGlnGluAlaAspTyrLeuSerGlySerGluAspProLeuGlu 526
26584 CCGGTCCA.....GGCTCATCGCTCCCGAGCG 26556
527 ArgSerPheLeuTyrCysAsnGlnAla.....ArgGlySerProAla 541
26555 AGA.....TGTGGAGAGACCGCGCTCAAGCTCTCCACCTCG 26518
541 gValAsnSerAspSerLysAlaLysProGlnSerValSerThrCysSerV 558
26517 CCGAGAAAGCCACCTCGGCGACGTAG.....GTCTGTCTCT 26481
558 alThrSerProAspAspMetSerThrLysSerAspSerPheGlnSer 574
26480 GTGCTCTCTCCGAGCGG...GGAGCGCGCGCTCTGCCCATTCACCGG 26434
575 LeuHisGlySerLeuProLeu..... 581
26433 TCGCGCGCACCTCGGTGTGACACACGCGCTCGGTCAACAGCTGCCA 26384
582 ...AspSerPheSerLysPheValAlaGlyGluArgAspCysProArgL 597
26383 GTAGACCTCCAGCTCTCGACCGCGCGCGCGGAGCGG..... 26347
597 euLeuLeuSerAlaLeuAlaGlnGluAspLeuAlaSerGluLeuLeuGly 613
26346CAGGCGCACCCAGATGATGCGATCGGTCTCTGTCGCGGC 26305
614LeuGlnGluAlaLeuGlyGluLysAlaAspLysAlaTrpAl 627
26304 AACGAGCTCTCCGCGACGTCGCGTCTCTGACGCTCCAGCGCTGGCT 26255
627 aGluAlaProSerLeu.....V 633
26254 CAGCAAGTATTCACACGCGCTCCAGGTCGATGTCGACGCTAGCG 26205
633 alLysAspSer.....SerLysProProPheSerLeuGluAsnHis 646
26204 TCGCGGACAGCGCATACCCAGCTCACCTGAAGCGCTTCGGGATCTCC 26155
647 SerAlaCysLeuAspSerValAlaLysSerAlaTrpProArg..... 660
26154 ACGGCATCAGGAGTTCGAGGCGCTCTCGCGAGCGCTCTGCGACGTC 26105
661ProGlyGluProGlnAlaLeuPro..... 668
26104 GAGCGTGCGCTGTCGTTAAAGCCATCACCGCGCGACACGCGCGCA 26055
669AspSerLeu.....Gln 572
26054 CAGCTCTAGAGCGCGGAGCGGTCTCCACACAGACGCGGTTCGCGCCAG 26005
673 LeuAspLysGlyAlaAsnAlaLysAspPheSerProGlyLeuPheGluAs 689
26004 CGTCTCAGAGTGGCAGGCGGAG.....GAGA 25976
689 pProSerValAlaPheAlaThr...ProAspProLysThrGly... 704
25975 GGCCTCTTGTGGCGAGTTACACGCGCATCCAGAGCGCGCTCGGCTCG 25926
705ProLeuSerPheGlyThrLysProThr..... 713
25925 CGTCCGCGCGCGCTCCGCGATCGGCGCATTCATCTGGATGACACG 25876
714LeuGlyValProAla...ProAspProT 722
25875 CGCTCGTGGCGGCGGTACGAGCAGCGCATTCATCCGCGCGCGCGCG 25826

722 hzThrAlaAlaPheAspCysPheProAspThrThr.....AlaAla 735
25825 ACTCGTGGCATTCGCCAGATCCCGATGCTCGTGTTCGCCGCGCTCG 25776
736 SerSerAlaAspSerAlaAsnProPheAlaTrpProGlnGluAsnLeuG1 752
25775 CCGTGGAGCCATCCCGCTCCGCCACAGCGCGCGATGCTCAGG 25726
752 yAspAlaCysProArgTyrPheGlyLeuHisProGlyGluLeuThrLysGlyL 769
25726 GCGCGAGCGCTTCGCTTCGCTGCG..... 25704
769 euGlnGlnGlyGlyLysAlaSerAspGlyLeuSerLysGlyAspThrHis 785
25703GGAGGT..... 25698
786 GluAlaSerAlaCysLeuGlyPheGlnGluGluAspProProGlyGluLys 802
25697CCGCCAGCG..... 25689
802 sValAlaSerLeuProGlyAspPheLysGlnGluGluValGlyValL 819
25689 25689
819 ysGluGluAlaGlyGlyLeuLeuGlnCysProGlnValAlaLysAlaAsp 835
25689 25689
836 ArgTrpLeuGluAspSerArgHisCys.CysSerThrAlaAspPheGlyA 852
25688CGTCCAAAGGCATTCCTCGCGGTAGCTGCCCTGGCGG 25648
852 sPLeuProLeuLeuProProThrSerArgLysGluAspLeuGluAlaGlu 868
25647 AT.....CGAGCGAGAGAC..... 25634
869 GluGluTyrSerSerLeuLeuLeuLysSerProGlnGluArgPr 885
25633GCCGAGCGCG 25623
885 ocLysMetGlnAspProLeuSerProLysAlaProLeuLeuLysG 902
25622 AGGAAACAG.....TAGC 25609
902 luGluValGluGluValLeuAspSerLysAlaGlyTrpGlySerProCys 918
25608 AAGAGGTCCAGCGGTCTCGCGGTAG...GGTGTGACAGCACCATGC 25562
919HisLeuSerGlyGlu.SerValIleLeuLeuGlyProThrValG 933
25561 CCCTCCACCTTGGGCGCAACACC..... 25537
933 yThrGluSerLysValGlnSerTrpPheGluSerSerLeuSerHisMet 949
25536 ..CGGCGGCGCGACGCGTCTGTGTGGCGACGCGCTCGTCGAGC 25489
950 LysPro.....glyGluGluGlyProAs 957
25488 AGACCGCGCGGTGCACTACCCCGCGCGGCGGTCTGACGCGCGCCAA 25439
957 pGlyGluArgAlaProGlyAspSerThrThrSerAspAlaSer..... 971
25438 GAGCGCGCGCATCGCTCGGCATCGGCCAGCTCGACGCGCGCGCGGTGA 25389
972LeuAlaGlnLys.ProAsnLysProAlaValProGluAl 984
25388 CCGCGCGCGCTCGGCTTCAGAGCGCTCGATCGCGCATGCGCGCGCGC 25339
984 pProIleAlaLysLys...GluProValProArgGlyLysSerLeuArg 1000
25338 ACCTCTGCGCGGTGATCTCGGCGCGCATTCCTCGCGTCCGCGCAATCCGTG 25289

1000 *errhrgrgtaHhshargyleuPro*..... 1008
 25288 *CCGGCTGATGACGACAGAGTGGCCCGCCGCGCTCCACCAACACC*CGG 25239
 1009 *GLuAGlAAspSerProCysargtAlaPr* 1018
 25238 *CACAGGAGAGCCAAAGGCGCCCAAGCCACCCGATCCCAAGTAAC*TCCCG 25189
 1018 *oVallenuProLysAspLeuLeuProGluSerCysThrGly*..... 1032
 25188 *T*..... *CCGCAAGAGAGACACGGCGCTGC* 25163
 1033 *ProProGlnGlyGlnMetGlnGly*..... *AlaGlyAlaPro* 1044
 25162 *GTTTCCCTCCGATGGGGCGGACACACCGCGCTGCGCGCGGGCCCT* 25113
 1045 *GlyArgGlyAlaSerGlnGlyLeuProArgMetCysThrArgSerLeu*h 1061
 25112 *GGC*..... *GAATGCGACACTGATCCTCGGCTCGGCGGA* 25078
 1061 *lAlaLeuSerGlnProArgThrProGlyPro*..... 1071
 25077 *AGCA*..... *GCTGCGCACAGCAGGCGCTCGAACCCTGCGGGCTGTC* 25037
 1072 *ProGlyLeuThrThrPro*..... 1078
 25036 *CTCCGGATTCAGAGTCCACAGAGCCCGCCCAAGAGCGGAGATGCTCAAG* 24987
 1079 *AlaProPaspLys*..... 1083
 24986 *CCGCGACCCGCGCCATATCCCAACAGCGCGGCTGACAGAGGGGCGCTCA* 24937
 1084 *LeuGlyGlyGlnArgAlaAla*Pr 1092
 24936 *GCGTCGCGCCACACCGAGCGAGCGCCCTCGGGTCAACATCCAGACCGGG* 24887
 1092 *he*..... *LysSerGlyLysArgAlaGlySerPro* 1103
 24886 *TGAAGCGGCGCCCGCTGCGCGAGCGCTGATACAGGCGAGACCGGCGCG* 24837
 1104 *Lys*..... *AlaAlaSerProSerAsp* 1112
 24836 *CGGAGAGATGGGTACTTTGGACACTCTTGGCGGCATGCCCCCGCTCC* 24787
 1112 *calAlaLeuProValAlaSerAspSerProMetGlyLysThrL* 1129
 24786 *ACGACGGCGTCCAGACCCACAGTACAGACACCCCGCTGCAGTGGTGG* 24737
 1129 *ysGlnThrAspSerProSerThrProGlyLysAspAlaArgSerMetLe* 1145
 24736 *GCCACCCA*... *GGGCGTGGGTGACCTGCTCGGCAACCGCG*... 24699
 1146 *LeuArgSerArgThrLysThrGlnGlnLeuPheHisSerLysArgArg* 1162
 24698 *AGGCTCGGCG* 24688
 1162 *gProSerGlnGlyArgLeuProAsnGlyArgAlaThrLys*..... 1175
 24687 *GGCCATGAGAGACAGCGCGACAGCAACTCCCGGAGCAAAAGCGCGCGC* 24638
 1176 *LysLeuLeuAspAsnSerHsLeuProAlaThrPheLys* 1188
 24637 *GAGCGCTCCCGACACTCCACCGCGGCGGACAGACAGACCGCGCGG* 24586
 1189 *ValSerSerSerProGlnLysGlnGlyArgValSerGlnArgAl* 1203
 24587 *ACCGAGGCTCCGCGAATCCA*..... *CGGATGAGAGAGCATC* 24550
 1203 *argGlyProLysProGlyAlaGlySerLysLeuSerAspArgPro*... 1218
 24549 *TGGGCGCAAGTCCACCGGTAAACACACTCGGCGGCAAGCGGAGCGGT* 24500
 1219 *LeuHsAlaLeuLysArgLysSerAlaPheMetAlaProValProThr* 1234

24459 GGCTCCGAGCCCTTCGG.....CCGCGCGCT... 24474
125 LysLYSArgAsnLeuValLeuArgHisGlySerSerSerSerSerAsnI 1251
12443CGATCCAACTACCCCTCCGCGCCACAGATAGTGGGACGCGC 24430
1251 AsnAlaMetGlyLysMetGlyArgArgArgGlyLeuArgValProPro. 1267
24429 ACCCGCGCGCCACGCCGCGGAGACG.....CCGCGCA 24395
1268ProSerSerArg.....GlyCysLeuLeuProArg...Lys 1278
24394 GCGACCGAGCGCGCGCGGCCGCCAAGCGTGCACAGCGCTCGAGCACCC 24345
1279 ProSerProProArgAla..MetAlaSerLeuProGlnSerSerHisSer 1294
24344 CCGGACGCTCCCGCGCGCGCGCCACACAGCCAGCGAGCGTGGCGTCC 24295
1294 ArgArgArgProProPheLeuProGlnAlaArgLeuSerAlaAlaPheGln 1311
24294 GCGTCCGCGGACGAGCGGTGGCACACGCCGACGAGCGTGGTGGCGGCC 24245
1311 LysAlaMetLysThrLysValLeuProProArgGlySerGlyArgLys 1327
24244 CACTTCGAGAACGTC.....CCGCGCGCGGTGCT..... 24213
1328 LeuGlnAlaLeValGlnLysLerHisSerProSerLeuLysLysPheAl 1344
24212GCAGCGCTTCACCCCGCGCGCAGAG 24187
1344 AcCysLysAlaProGlyAlaSerProGlyLysProLeuSerProSerLeus 1361
24186 C.....GCACCGCTCCCGCGAGTGCAGCCACCGATACCCGCGCGCT 24143
1361 ArgAspLysAspArgGlyLeuLysGlyAlaGlyLysProValGlyVal 1377
24142 CAGGTGCTCGCGACACACCTCCCGCGTGGTCTCACCACG..... 24059
1378 GlnGlnLysLeuValAsnValGlyThrGlyGlnLysLeuProThrSerGI 1394
24098AAACGCTGGCGCGCGGT.....ACATCAC 24073
1394 ValAspProLeuCysArgAsnProThrAsnArgSerLeuLysGlyLysL 1411
24072 GACCGCGCCACCGCGCGCACTCTCCACACATCGGTCCATCGAGCGCA 24023
1411 eMetAsnSerLysLysLeuSerSerThrAspCysPheLysThrGlnAla 1427
24022 GCGAGAGCGTGGAGAAATCGAACCGCGTGTGCGCGCGCGCGCG 23973
1428 PheHisSerProGlnAlaLeuGlnProGlyGlyThrAlaLeuAlaProLy 1444
23972 CGAACCCCGCGCGAGTCCCTCGACG.....CTTGCTCACG 23935
1444 sLysArgSerArgLysGlyArgAlaGlyAlaHisGlyLeuSerLysGlyP 1461
23934 CCGCGGATCACACCT.....GTCGCGCGCATTCAGCCGCGCATGCA 23891
1461 rLeuGlnLysArg.....ProTyrLeuGlyProAla 1471
23890 CACCGAGCGCGCTGGCGGCGCACCGCGCGCCACTCGCGCTCGCGG 23844
1472 LeuLeuLeuThrProArgAspArgAlaSerGlyThrGlnLysAlaSerGI 1488
23840 CTCGAGAGGACACACAGCGCGCGCGAGAGCGCTCATACAGCC... 23793
1488 AspAsnSerGlyGlyGlyLysLysProLysMetGlnGlnLeuGlyP 1505
23792GCCGCGCGCGCGACGACGACCGCATTC 23755
1505 rAlaSerGlnProProGlnLysArgProCysGlnProGlnThrArgAla 1521

```
23758 CAGCGAGAACACCCCGCCACGACGCG...CCACAGCTCCCGATGC 23712
1522 GlnLysGlnProGly.....HisThrAspTyrSerSerty 1533
|||||
23711 TATGCCAACAGAGCTCGGCTACGCCCCACGACCGCCACAGCGCC 23662
|||||
1533 rSerLysArgLysLeuThrArgGlyArgAlaLysAsnThrSers 1550
|||||
23661 GTCAGCGGTACTCCACCGGAGAGCGCGGTGGTGAACCGGTCT. 23613
|||||
1550 erProCysLysGlyArgAlaLysArgArgGlnGlnValLeuPro 1566
|||||
23612 .....GGTCGAGACACAGACTCGGCGCTCC.....CCG 23583
|||||
1567 LeuAspProAlaGluProGluLeuLeuArgLeuLysTyrLysSerS 1583
|||||
23582 CTTCCCGCCACATCACT.....CGCGCAGCGGGGTCACG 23545
|||||
1593 sArgLeuArgSerAspSerArgThrProAlaPheSerProPheVal 1600
|||||
23544 TCCTGGTCAACAGCGCCACGACCGCGGTGCAACCGCTCC..... 23505
|||||
1600 alGluLysArgAspAlaPheThrThrIleCysThrValValAsn 1616
|||||
23504 .....GGACGCTGGCCACGCGCGCAAGCGCC 23475
|||||
1617 GlyAspAlaPro.....LysProHisArg.....LysProSe 1627
|||||
23474 GGCCATGTCGGCGCTCTGGCGCGCTGTCGGTGAACAGGACCGC 23425
|||||
1627 rSerSerAlaSerSerSerSerSerSerPheSerLeuAspAla 1644
|||||
23424 TTGCGCGGAGGAGCTCGCGATGCAGC..... 23397
|||||
1644 laGlyAlaSerLeuAlaThrLeuProGlyGlySerIleLeuGln 1660
|||||
23396 .....CGCGCGCGCGCGCGCTGCTGCTGCGCGCGCG 23361
|||||
1661 ProSerLeuProLeuSerSerThrMetHisLeuGlyProValVal 1677
|||||
23360 CGGAGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCG 23311
|||||
1677 sAlaLeuSerThrSerCysLeuValCysCysLeuCysGlnAsnPro 1693
|||||
23310 CGGTGGCTCATCGCGCTC.....GCGTCTCGCGCGCG 23276
|||||
1694 .....AsnPheLysAspLeuGlyAspLeuCysGly..... 1703
|||||
23275 GAACGCGACGTCGCGCGCGCGCGCTGCGGTGCGATGTCAGTGT 23226
|||||
1704 .....ProTyrTyrProGluHisCysLeuProLysLysLysPro 1717
|||||
23225 GCAGCGCGCGCG.....CTGCGCATCGAGCGCGCGCGCTGTC 23182
|||||
1717 sLeuLysGlyLysValArgProGluGlyThrCysGluGluAlaSer 1734
|||||
23181 GACAGCACCAAGAGTCCGCGAGCGCTCGCGCGCGCGCGCGCG 23132
|||||
1734 roLeuGluArgThrLeuLysGlyProGlu..... 1743
|||||
23131 CACCGCGCGCGCTCTTCACACACACATGCGGTTCCTGCTCATCC 23082
|||||
1744 .....CysAlaAlaAlaAl 1748
|||||
23081 CGAAGAGCTCACCCCGGAGCGCGCGGTGCGGTGCGCGCGCGCG 23032
|||||
1748 aThrAlaClyLysProProArg 1755
|||||
23031 ACCGGCTCGTCCCAACGCA 23010
|||||
```

seq_name: /cgn2_6/ptdata/2/ina/6B_comb.seq:US-09-443-501A-2

seq_documentation_block:

```
; Sequence 2, Application US/09443501A
; Patent No. 630342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; Epithelone and Epithelone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443.501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
; US-09-443-501A-2
```

alignment_scores:
Quality: 462.50 Length: 2116
Ratio: 0.518 Gaps: 119
Percent Similarity: 42.202 Percent Identity: 22.590

alignment_block:
US-09-508-821B-6 x US-09-443-501A-2/rev ..

Align seg 1/1 to reverse of: US-09-443-501A-2 from: 1 to: 71989

```
33 ProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAs 49
|||||
22992 CCTTCCGCTCAGGTGCTCA...CCAGCGAAGACGCTTGGCGCGGTAC 22946
|||||
49 pTyrTrAsnProGlnProTyrProSerTyrGluGlyAlaGlyThrP 66
|||||
22945 G.....TCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 22902
|||||
66 roSerGlyThr.....AlaAlaAlaAlaAla 75
|||||
22901 CATCAGCGCGAGTGGATGCGTGGAGACATGCGCGCTTGGTGGCA 22852
|||||
76 AspLysTyrHisArgGlySerLysAlaLeuProThrGlnGlnGlyLeu 92
|||||
22851 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22802
|||||
92 nglyArg...ProAlaPheProGlyTyrGlyValGlnAspSerPro 108
|||||
22801 CCGCGGATCACCACTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22752
|||||
108 yrProGlyArgTyrAlaGlyGluGluSerLeuGlnAlaTrpGlyAla 124
|||||
22751 CGCGG.....CGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 22729
|||||
125 GluProProPro.....GlnProGln 132
|||||
22728 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22679
|||||
132 nProLeuProAlaGlyValAla.....LysTyrAspGluAsnL 145
|||||
```


20184 C.....CTCATGTTCCGGGGTGCCTGGAGGCCATCCCGCCGTCC 20141
| | | | | | | | | | : : : : :
892 ePrProlyAlaPro..... 896
| | | | | | | | | | : : : : :
20140 GCCACAGGCCCCAGCGCATGCTCAGGGCGGAGCCCCCTGCTTCGAGC 20091
| | | | | | | | | | : : : : :
897LeulleCys..... 899
| | | | | | | | | | : : : : :
20090 GAGGTCTGCCAGCGCTCCAAAAAGGCATTGGCTCGCGTAGTGCCCT 20041
| | | | | | | | | | : : : : :
900 eaThrysluGluy 904
| | | | | | | | | | : : : : :
20040 GCCGATGACCGAGAGCGCCGACCGCCGAGAAAACATGACERAGAG 19991
904 aiglulnValLeuAspSerLyssalaGlyTrpGlyserProCys..... 918
| | | | | | | | | | : : : : :
19990 TCACGGCTGCTCGGGTAAG...GGTGTGACACCATGCCCCCCTC 19944
| | | | | | | | | | : : : : :
919 HisLeuSerGlyGlu_SerVallIleLeuLeuGlyProThrValGlyThrG 935
| | | | | | | | | | : : : : :
19943 CACCTTGGGGCGCAACCC.....CGGG 19921
| | | | | | | | | | : : : : :
935 luSerLyssalInSertirPhegluSerLeuSerHisMetLysPro 951
| | | | | | | | | | : : : : :
19920 CGAGCCAGCCGGGTCTGTGTGGCCAGGACCGCTGCTCGAGCACACC 19871
| | | | | | | | | | : : : : :
952GlyGluGluProaspGlycl 959
| | | | | | | | | | : : : : :
19870 GGCGGTGCACGACCCCGGAGCGGGCTCGAGCGGCCGCAAGAGCG 19821
| | | | | | | | | | : : : : :
959 uArgAlaProGlyYasSerThrThrAspAlaSer..... 971
| | | | | | | | | | : : : : :
19820 CGCCATGCTTCGCATCGGCCACGTGACCGCCGCGGACCGCG 19771
| | | | | | | | | | : : : : :
972LeuAlaGlnLys_ProasnLysProAlaValProGluAlaProIl 986
| | | | | | | | | | : : : : :
19770 CCCTCGTGGCTCGACGGCTCGATCGCCGCATCGCGCGCGACCTCT 19721
| | | | | | | | | | : : : : :
986 eaLatLysLys...GluProValProArgGlyLysSerLeuArgSerArgA 1002
| | | | | | | | | | : : : : :
19720 GGCGCTGATCTCGGCCCATCTCTCGGTGGGCANTCGTGGCGGT 19671
| | | | | | | | | | : : : : :
1002 rgValHIsArgGlyLeuPro..... 1008
| | | | | | | | | | : : : : :
19670 GATGACACAAAGTGCCTCGCGCGCTCCACCAACACGCGCAACGA 19621
| | | | | | | | | | : : : : :
1009GluAlaGluAspSerProCysArgAlaProValle 1020
| | | | | | | | | | : : : : :
19620 GGAGCCGAGGGCCGACCCGACCTCCACNACTACTCCCT.... 19576
| | | | | | | | | | : : : : :
1020 uProLysAspLeuLeuProGluSerCysThrGly.....P 1033
| | | | | | | | | | : : : : :
19575CGCAGACAGGACGCGGCTGGCTTCC 19545
| | | | | | | | | | : : : : :
1033 roProGlnGlyInMetGlu.....AlaGlyAlaProGlyArg 1046
| | | | | | | | | | : : : : :
19544 CTCCTGGTGGGGCCGACAGCGCGGTGGCGCGCGCCCTGGC... 19498
| | | | | | | | | | : : : : :
1047 GlyAlaSerGluGlyLeuProArgMetCysThrArgSerLeuThrAlaLe 1063
| | | | | | | | | | : : : : :
19497GGAATGCCACTGATCTCTGGGTCTGGCGGAAGCA.. 19462
| | | | | | | | | | : : : : :
1063 userGluProArgThrProGlyPro..... 1071
| | | | | | | | | | : : : : :
19461GCTGGCCACAGGSCCTCGAOTCCGTGGGTCTCTCCGG 19419
| | | | | | | | | | : : : : :
1072 ..ProGlyLeuThrThrPro..... 1078
| | | | | | | | | | : : : : :
19418 ATCCAGTTCACAGGCCGCCGCCAGAGCGGGGTGCTCCAGCGCGCA 19369
| | | | | | | | | | : : : : :
1079AlaProAspLys..... 1083
| | | | | | | | | | : : : : :

19368 CCGGCGCCATACCCGACGCGCCCTGACAGGCGGACAGGCTAGGCTCG 19319
1084LeuGlyGlyLysGlnArgAlaIlePheLys 1094
19318 CCGCCACCGGTGACAGGCCCTCGGGTCACAGATCCAGAGCGGGGTAGCG 19269
1094 ergGlyLysArgValGlyLysProSerProLysAlaIleSerProSer 1110
19268 CCGGCCCGG.....TGCACACCGCTGATACAGCGAGACCGCGGCG 19228
1111As 1111
19227 CCGGGGACAGATGGGTACTTCGCACTCTTCGCGCCATGCCCCCGCG 19178
1111 nProAlaIleProValAlaSerAspSerProMetGlySerLysT 1128
19177 TCCACGACGCGGCTCCAGACCCGACAGGTACAGACCCCGCGACGCTT 19128
1128 hTlyGlnThrAspSerProSerThrProGlyLysAspGlnArgSerMet 1144
19127 GCGGCGACCGA...GGGCTGGGTACCTGCTGCGGACGCGG..... 19087
1145 IleuArgSerArgThrLysThrGlnGlnIlePheHisSerLysArg 1161
19086AGGCTCG 19079
1161 gArgProSerGlnGlyArgIleProAsnCysArgAlaThrLys..... 1175
19078 GCGGCGCATGAGACGCGACGACGACGACGACGACGACGACGACGACG 19029
1176LysIleuAspAsnSerHisIleuProAlaThrPhe 1187
19028 CCGGCGCGCTCCCGACTCCACCCGCGTGGCGACGACGACGACGACG 18979
1188 Lys.....ValSerSerProGlnLysGlnGlyArgValSerGln 1202
18978 CCGACCGGGCTCGCGCGATCA.....CGATGACGACGACG 18941
1202 gAlaArgValProLysProGlyAlaGlySerLysSerAspArgPro 1218
18940 ATCTGCGCGCATGTCACCGGTTAGAACCACTGCGCCAGCATGCGGCGG 18891
1219LeuHisAlaIleuLysArgLysSerLysPheMetAlaProVal 1233
18890 CCGGCTCGGACGCTTGG.....CCGCGCGCT 18862
1234 ThrLysLysArgAsnIleuValIleuArgHisGlySerSerSerSer 1250
18861CGATCCAGTACCGCTGCGCGCGACGACGATGCGGACG 18821
1250 nAlaSerLysMetGlyLysMetGlyArgArgArgGlyLeuArgValPro 1267
18820 GGCACCGCGCGCCGACGCGTGGGAGAGACG..... 18790
1267 roProSerSer.....ArgGlyCysLeuLeuPro 1276
18789 ..CCGCGACGTACCGACGCGCGACGCGCCACACGCTCGCCAGCGCT 18742
1277 Arg...LysProSerProProArgAla...MetAlaSerLeuProGln 1291
18741 CCGACGACCCCGACGCTCTCTGCGCGCGCGCGACGACGACGACGAC 18692
1291 rSerHisProArgArgProProPheLeuProGlnAlaArgLeuSerLys 1308
18691 GTCGCTCGGCTCGCGGACGACGACGACGACGACGACGACGACGACG 18642
1308 laphGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArg 1324
18641 CT...TCGGGCGCATTCGAGAGACGTC.....CCGGCGGCTTCGT 18601
1325 GlyLeuLysLeuGlnAlaIleValGlnLysIleThrSerProSerLeu 1341
18600CGACGCGCTTCACCGCG 18584

1341 sLysPheLysCysLysAlaProGlyAlaSerProGlyLysnProLeuSer 1358
18583 TCCGGGAAGC.....GACGCGCTCCCGGACGTCGCGGCCGACGTAAGC 18540
1358 roSerLeuSerAspLysAspArgGlyLeuLysGlyAlaGlyLysPro 1374
18539 CCGCGCGCTACGCTCGGTACACGCTTCGCGCGACGCTTCGACCA 18490
1375 ValGlyValGlnGlnGlyLeuValAsnValGlyThrGlyGlnLysLeu 1391
18489 GCG.....AAAGCTGGCGCGCGT..... 18469
1391 oThrSerGlyAlaAspProLeuCysArgAsnProThrAsnArgSerLeu 1408
18468 .AGCTACCGACGCGCGCGACCGCGCGGACGCTTCGACGACGCTTCAT 18420
1408 yseGlyLysLeuMetAsnSerLysLysLeuSerThrAspCysPheLys 1424
18419 CACCGCGACGAGAACCGCTGCGACATGACGCGCTTCGTCGCGACG 18370
1425 ThrGlnAlaPheThrSerProGlnAlaLeuGlnProGlyThrAla 1441
18369 CCGCGCGCGGACCGCGCGCGGACGCGCTTCGACG.....CT 18332
1441 uAlaProLysLysArgSerArgLysGlyArgAlaGlyAlaHis..... 1455
18331 TGCTCCAGCGCGGATCACACT.....GCTCGCGCGCTTCGACGCG 18288
1456GlyLeuSerLysGlyProLeuGlnLysArgProLysLeu 1468
18287 CCGGACGACGACGACGCGCGCGGCGCGGCGCGCGCGCGCGCGCG 18238
1469 GlyProAlaLeuLeuThrProArgAspArgAlaSerGlyThrGln 1485
18237 CCGCGCGCGCTCCGCGACGACGACGCGCGCGCGCGCGCGCGCG 18188
1485 yAlaSerGlnAspAsnSerGlyGlyGlyLysLysProLysMetGln 1502
18187 ATGACG.....GCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18156
1502 IuLeuGlyProAlaSerGlnProProGlnGlyArgProCysGlnPro 1518
18155 CCGATCTCCAGCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 18109
1519 ThrArgAlaGlnLysGlnProGly.....HisThrAsnTy 1530
18108 CCGGATGCTATGCCAACCGAGGCTCGGCTTCAGCCCGACGACGCG 18059
1530 rSerSerLysLysArgLysArgLeuThrArgGlyArgAlaLysAsn 1547
18058 CAGAGCGCGCTGACGCGGTCTCCACCGCGAGAGCGCGGTGGGTGA 18009
1547 hThrSerSerProCysLysGlyArgAlaLysArgArgGlnGln 1563
18008 CCGCGCTGCTCCAGCA.....ACAGACGCTCGCGCGCTCC..... 17974
1564 ValLeuProLeuAspProAlaGlnProGlnIleArgLeuLysIleSe 1580
17973CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17942
1580 rSerCysLysArgLeuArgSerAspSerArgThrProAlaPheSerPro 1597
17941 CCGTCCAGCTCCCGGTGAAACGCGACGCGACGCGCGCGCGCGCGCG 17893
1597 heValArgValGlnLysArgAspAlaPheThrThrLysThrVal 1613
17892GGAAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 17872
1614 AsnSerProGlyAspAlaPro.....LysProHisArg..... 1624
17871 AGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 17822

```

1625 .LysProSerSerSerAlaSerSerSerSerSerSerSerSerSer 1641
17821 AACGCCACCTGGCGCGGAGAGCTCGCATGACG..... 17785
1641 euAspAlaAlaGlyAlaSerLeuAlaThrLeuProGlySerIleLeu 1657
17784 .....GCGCGCGCGCGCGCGCTGCGCCCT 17758
1658 GlnProArgProSerLeuProLeuSerSerThrMetHisLeuGlyProVa 1674
17757 GCGCACGCGCGGAGCGCGCGCGCGCGCGCTCGCGGAGCTCAGCGGC 17708
1674 lValSerIysAlaLeuSerThrSerCysLeuValCysCysLeuCysAla 1691
17707 ACCGGAGCGCGTGGGTATCGCGCTC.....CGGTCT 17673
1691 snProAla.....AsnPhelYasPleuGlyAspLeuCys 1702
17672 CGCCAGGTGAGCGCACGCTCCCGGAGCGGAGCTCGGGTGCTATGCCA 17623
1703 Gly.....ProTyrTyrProGlnHisCysLeuProLysLy 1714
17622 GTGTCTCGCGAGCGCGCG.....CCTGCGCTCGAGCGCGCTCG 17579
1714 slsProLysLeuLysLysValArgProGlnGlyThrCysGlnGluA 1731
17578 CTCTTGGCGGACGACGACCAAGCTCCGCGGAGCGCTCGCGCGCGG 17529
1731 lAserLeuProLeuGluArgThrLeuLysGlyProGlu..... 1743
17528 CCACAGCTCCACCGCGCGCGCTCTCCAGCACACATGCGGTTCGTT 17479
1744 .....CysAl 1745
17478 CGTCATCCCGAAGAGCTCACCCCGGAGCGGAGCTCGGTGCGC 17429
1745 alaAlaAlaThrAlaGlyProProArg 1755
17428 GGCACGCGCGCGCTCGGTGCGCGAGCGGA 17398

```

```

seq_name: /cgn2_6/prodata/2/ins/6B_COMB.seq:US-09-105-537-1

```

seq_documentation_block:

```

; Sequence 1, Application US/09105537A
; Patent No. 6263202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

```

alignment_scores:

```

Quality: 454.50      Length: 2052
Ratio: 0.521        Gaps: 117
Percent Similarity: 42.544      Percent Identity: 23.635

```

alignment_block:

```

US-09-508-821B-6 x US-09-105-537-1
Align seg 1/1 to: US-09-105-537-1 from: 1 to: 15872

```

```

31 ArgGlnProSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAl 47
639 AGAGCCCGCTGACGAGAGCGCTTCGGTGACCTCCCGCG..... 680
47 aLysAspTyrTyrAsnProGlnProTyrProSerTyrGlnGlyGlyAla 64
681 .....ACGGCAGCGCTACACCTTCGAGCGCGGG..... 710
64 lYThrProSerGlyThrAlaAlaAlaValAlaAlaAspLysTyrHisArg 80
711 .....CCACGATTCCTCGGGGCGAGGCGCGGAGTCG..... 746
81 GlySerLysAlaLeuProThrGlnGlnGlyLeuGlnLysArgProAlaPh 97
746 ..... 746
97 eProGlyTyrGlyValGlnAspSerSerProTyrProGlyArgTyrAla 114
747 .....TCGTACTCAAGCGCTCTCCGCGCGCTCGCGAGCGGACGCTGT 792
114 lYGlnGluSerLeuGlnAlaThrGlyAlaProGlnPro..... 126
793 CCACGGCGTCA.....TCGGCGCGAGCGCGGCAACACGAGCG 830
127 ProProGlnProGlnProLeuProAlaGlyValAlaLysTyrAspGl 143
831 GAGCACCGCGGCTCAGCTCGCGCGAGG..... 863
143 uAsnLeuMetLysLysThrAlaValProProSerArgGlnTyrAlaGlu 160
864 .....CGGCCAGGAGAAGTGTCTGCGCGA 888
160 lN.....GlyAlaGlnValProPheArgThrHisSer 170
889 GCGTACCGGAGAGCGCGCTGGACCCCGCGCGCTCCAGTACGTCGAAC 938
171 LeuHisValGlnGlnProPro.....ProGlnGlnProLeuAl 184
939 TCCACGACCGGAGAACCGCTCGGGGACCCATCGAGCGCGCGCTC 988
184 aTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerProL 201
989 GCGCGCGCTCTCGGTCTGGCGC.....GCCGCGC 1017
201 euProPheProGlnGlyThrHisPheProGlnHis.....Ser 213
1018 GAGCAGACCCCTCTCTCGGCTCGGCCAAGACGACGTCGGGCACTCG 1067
214 GlnSerPheProThrSerThrTyrSer.....SerSerValGlnGlyG 229
1068 AAGCGCGCGCGGATCTCTCGGCTCATCAAGACGCTCTCTCGGCTCG 1117
229 rGlyGlnGlyAlaHisSerTyrLysSerCysThrAlaProThrAlaGln 246
1118 GCGCGCGGATCCGCGAGCTCAACTTCGAGCGCGCGCGCGGACAT 1167
246 rHisAspArgProLeu.....ThrAlaSerSerSerLe 257
1168 CCCGCTCGACACCTCGGCTCGAGCTCGCGCGCGCTCGGGAGTGGC 1217
257 uAlaProGlyGlnArgVal.....GlnAsnLeuHisAlaTyrGlnSerGly 273
1218 GCGACCGCGGACCGGAACTCTCTCGGCGCTCAGCTCTGTCGCGATGG 1267
273 rGLeuSerTyrAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 289
1268 GG.....CACCAAGCGCGAGCTCTCTCAGGAGAGCGCGCGCG 1305
290 GlnGlnAlaLeuGlnSerArgHisHisAlaGlnGlnThrLeuHisTyrGl 306
1306 CCAGCGCGCGAGACGCGCGCTCGATGAGGAGACCGCGCGCGCGCG 1355
306 nAsnLeuAlaLysTyrGlnHisTyrGlyGlnGlnGlnGlnGlnGlnGln 323

```

[illegible]

2184 CGCAGCGCTTCACGTCGGCGTCCATGAGAGGCAGATGTGTCGAGCAATTCCGC 2233
586 LysPheValAlaGlyGluArgSpCyS ProGLeuLeuLeuSe 600
2234 CGG.....GTGCTGGAAAGCGCTGGACTTTCCAGAAGCCCCCGCTTC..... 2272
600 rAlaLeuAlaIngInuSprLeuAlaSerGluIleLeuGlnGluIngLa 617
2273CACGTGCTGTCCACGGGTGAAGGAGGAGGCTGCTGTGTA 2306
617 lAlaGlyGluLysAlaAspLysAlaTrpAlaTrpAlaGlnAlaProSerLeuAl 633
2307 CAGCGGCCCA.....TGAGCCAGT...CCGAGTACTGG 2338
633 LysAspSerSerLysProProPheSerLeuGlnAsnHisSerAlaCyste 650
2339 GTGGACCGAAGTCCGAGAGCCCGTAGCCCTC.....CT 2370
650 uAspSerValAlaLysSerAlaTrpProArgProGlyGluProGluAla 667
2372 CGAGCCGCTA.....CGAGCCCGAGGAATCGGGCG 2402
667 euProAspSerLeuGlnLeuAspLysGlyGlyGlnAlaLysAspPheSer 683
2403 CCGAGACCTTCTCTGAAGCTGTGTCCCGAGAGGGGTCTGCTCCGCGATGGCG 2452
684 ProGlyLeuPheGlnAspProSerAlaLysAlaThr 696
2455 GCGGACCTCGCTACGAGAACAGAGAGCGCCGACCGCGCTTCGCGCCCTGGG 2507
697ProAspProLysLysThrThrGlyProLeu..... 706
2503 CAAGAGCGCCCGAGAGCCGACGCTGTCTGTGGCGGACATCACACAGCTCT 2552
707SerPheGlyThyLysPro 712
2553 TGTGCGGGGACAAGAGTGTGATGAGACCAGCCGCGAGAGGAGACCGCGC 2602
713 ThrLeuGlyValProAlaProAspProThrThrAlaAlaPheAsp 727
2603 AGGTGACGAGGTGCGCCCGCCG.....ACGTACGCTTCCACGCGCA 2643
728CysPheProAspThrThrAlaAlaSerSerAlaAspSerAla 742
2644 ACGGCACTAGTGTGAC.....GGGCGCGCGAGAGCGCGG 2678
742 snProPheAlaTrpProGlnGlnAsnLeuGlyAspAlaCys...ProArg 757
2679 GCGCGCTCAAGCGGGGCGCATGGGACCGGTGGGGGCAAGCCGCCGCC 2722
758 TrpGlyLeuHisProGlyGluLeuThrLysGlyLeuGlnGlnGlyGly 774
2723 GGCGGTGTGACGTCCGGGAG.....GGCAGAGGCGAAGGAGCGCGCGG 2775
774 sAlaSerAspGlyIleSerLysGlyAspThrHisGlnAlaSerAlaCys 790
2776 TGCGGAGTGGCGGTATGCGCGGAGCTCGCCAAAGAGACGAGCGCTGG 2822
791 LeuGlyPheGlnGlnAspProProGlyGluLysValAlaSerLeuPr 807
2826GCCAAAGATGTGGCGGCGCTCTCGAGTACAGAGACCGGACG 2866
807 obLysAr.PheYsGlnGlnGlnValGlyGlyValLysGlnGlnAlaGly 823
2867 GCGGTGCAATCGGCTGTACCTTCAGAA.....GTGGG 2901
824 GlyLeuLeuGlnCysProValAlaLysAlaAspArgTrpLeuGlnAs 840
2902CTTGACCTCCGTCATGTCGGTGAAGCT.....GGGAAAG 2936
840 pSerArgHisCysCysSerThrAlaAspPheGlnGlyAspLeuProLeuLeuAr 857
2937 GCCTGTGACACACAGGAGATCCGCTGCGCCAGCGAGATCTGTGTGAC 2986

857 ro.ProThrSerArgIysGluAspLeuAlaGluGlnGluLysr-SerSe 873
||| ||||| ||| ||||| |||||
2987 CACCCGAGCGCGCGGCCCTGGCGGCCAC 3016

873 rLeuCysGlnLeuLeuLysrProGlnArgProoLysMetGlnAspp 890
||| :||:||||| :||: :||: :||: |||||
3017 .CTGGCGCACCTGCTACC GCGCGCAGCGGCAGACCGATCGCGCAGC 3065

890 roLeuSerProLysAlaProLeuLecYstHrylsGluValGluGln 906
roLeu: ||| :||: ||||| ||| ||||| :||: :||: :||:
3086 GGATACCGCGCGACCCCGCGGACACACCGCGACCATCGCGATC 3115

907 Val..... 907
:||:

3116 ATCGCGATGGCTGCCGCTACCCGGCGCGTCACCTCCCCGAGGACT 3165

908LeuaspSerL 911
:||:

3166 GTGGCGGTCTCGCCGAGGGGCGGACGGCGTCTCGGGGCTCCCAACG 3215

911 ysAlaGlyTrpGlySerProCysHisLeuSerGlyGluSerValIleu 927
||||| :||: :||: :||: :||: :||: :||:
3216 ACCGCGGCTGGGACGAGGACCTCTCGAC 3244

928 LeuGlyProThrValGlyThrCluSerLysValGln..... 939
||| :||: :||: :||: :||: :||: :||:
3245 GCACGACCCCGACCGACCGGCAAGACTCGGTCGCGAGGCGGATTCT 3294

940SerTrpHecluaSer..LeuSerHisMetLysPro... 951
:||: :||: :||: :||: :||: :||:
3295 GCACGACGCGCGCTGTTCGACGCGCGCTCTTCGGGATATCCCGCGC 3344

952GlyGluGluCluProaspGlyGluArgAlaProGlyAspSerThr 967
:||: :||: :||: :||: :||: :||: :||:
3345 AGGCGCTCGGATGACCGCGCAGACGCGGTCTCTCGASAGCGCATGG 3394

967 hrSerAspAlaSerLeuAlaGlnLysProAsnLysPro....AlaValPro 982
:||: :||: :||: :||: :||: :||: :||:
3395 GAGGCGCTGGAGCGCAGGCGTGCACCGCGAGGCTCAGGCGACGCG 3444

983 GluAlaProIleAlaLysLys...GluProValProArgGlyLysSerLe 998
:||: :||: :||: :||: :||: :||: :||:
3445 GAGCGCGCTCTCTGTCGGCGCCACCGCCTTGAGTACGCGCGCGCATGC 3494

998 uArgSerArgValHisArgGlyLeuProGluAlaGluasp..... 1012
||| ||||| :||: :||: ||||| :||: :||: :||:
3495 ACAGCGCGCGCAGGCGTCTGAGGCGCACCTCTGACCGGACGACGCGC 3544

1013SerProCysArgAlaProValLeuProLysAspLeuLeu 1025
:||: :||: :||: :||: :||: :||: :||:
3545 AGGCTGATGTCGGCGCATCGCTACACGCTCGGCTCAC..... 3585

1026 LeuProGluSerCysThrGlyProProGlnGlyMetGluGlyAla... 1041

3586CGGTCTCGGTCACCGTCGACAGCGCGTCTCGCTCGCTCGCTG 3629

1042GlyAlaProGlyArg.....GlyAlaSerGlyLeuProA 1054
:||: :||: :||: :||: :||: :||: :||:
3630 TCGGCTGCACCTGGCGCTCGTTCGTCGGCGACGGGCGAGTCGAGCTC 3679

1054 rgMetCysThrArgSer..... 1059
|| | |||||
3680 CGCTCGCGCGGAGCGACCGTCATCTGACACAGCGGCATGTTCTGCGA 3729

1080 LeuThrAlaLeuSerGluProArgThrProGlyPro...ProGlyLeuthr 1075
:||: :||: :||: :||: :||: :||: :||:
3730 GTTCTCGCGCAGCGGCTCGCGCGCGCGCGCTCCA.....AGG 3773

1076 ThrThrProAlaProProAspLysLeuGlyLysGlnArgAlaAlaph 1092
:||: :||: :||: :||: :||: :||: :||:
3774 CCTTCTCGACTCGCGACGACACCTCTTGCGCGCGAGGCGTGCCTC 3823

1092 eIysSerGlyLysArgValGlyLysProSerProLysAlaAlaSerP 1109
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
3824 TCCTGTTCGGAGC...GGTCTCCGACCGCGACGCACGCCCAACC 3867
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
1109 roSerAsnPro..... 1112
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
3868 GTCTCTGGCGGTATCGGGGCGACGGCGGTACACAGACGGCGCTCCA 3917
1113AlaAlaLeuProValAlaSerAspSer..... 1122
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
3918 ACGGGTCTACGCCCCCAACGCCCGCTCCACGACGCGGTATCCGACAG 3967
1123ProMetGlySerThrLysGlnThrHisSerProSer..... 1135
3968 CCCCCTGGCCGACCGGGGCTCACCCGGCGACGCTGACGGCTGGAGGC 4017
1136ThrProLysAspGlnArgSer 1143
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4018 GCACGGTAGCGGTATCCGGCTGGCGACCCCATCGAGCCGAGGGATCC 4067
1144 MetIle.LeuArg.....SerArgThrLysT 1152
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4068 TCGGCACCTACGGCGGGGACCGGGGCGAGGGCGCTCGCTCAGCTCGCC 4117
1152 hrdGInGuilePheHisSerLysArgArgProSerGluGlyArgLeu 1168
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4118 TCGCTGAAGTGCAACAT.....CGGCCACGGCAGCGCGCGCC 4155
1169 ProAsnCysArgAla.ThrLysLysLeuLeuAspAsnSerHis....L 1183
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4195 GGGCGTGGCGGGCTCATCAAGAATGGTCTCGCATCGCCACGGCGTCC 4205
1183 euProAlaThrPheLysVal.SerSer...SerProGlnLysGluGlyAr 1198
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4206 TGCCAGGAGCTCCAGTGGACCGGGCCACACCGCGCTGCATGGGAG 4255
1198 gValSerGlnArgAlaArgValProLysProGlyLysAlaLysSerLysLeus 1215
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4256 CCGGCGGGCTCGAGCTCTCACGAGAGCGAGCGAGTGGCGGAGA...A 4301
1215 erAspArgProLeuHisAlaLeuLysArgLys.....SerAlaPheMet 1229
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4302CGGCGCGCGCGCGGGCGGCGATCTCTCTCGGCTCAGCG 4346
1230 AlaProValProThrLysLysArgAsnLeuValLeuArgHsigLy...Se 1245
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4347 GCACAACGCCCATCGTGTGCAAGAGCCCGCGAAAGCGGGGAGCGG 4396
1245 rSerSerSerSerAsnAlaSerAlaMetGlyGluMetGlyArgArgArg 1262
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4397 CGGTTCACCAACCGCCCGGAGACGGGGAAGCGGGGAGCGG... 4442
1262 lYleuArgValProProProSerSerArgGlycysLeuLeuProArgLys 1278
|||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4443CGGACACCGCCACCAACGAGC.....CGGCGCGGCTCG 4478
1279 ProSerPro...ProArgAlaMetAlaSerLeuProGlnSerSerHisPr 1294
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4479 CGCTCCCGCAACCGCTAACGG.....CCCGCTGTGTGTCTCC 4516
1294 oArg.....ArgProPheLeuProGlnAlaArgLeuSerAlaAlap 1309
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4517 CGGGGGACCGCGCGCTGGCGCCCGAGCGCTGGCTGGCGACTT 4566
1309 heGlnGlyAlaMetLysThrLysValLeuProProArgLysGlyArgLy 1325
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4567 CCTCGAGCGCCGACCGGAGCTCACCGTCCGCGACCTCGGACGCTGCTGT 4616
1326 LeuLysLeuGlnAlaIleValGlnLysIleThrSerProSerLeuLysLy 1342
|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
4617CGCGCGTGAACGCC 4630
1342 sPheAlaCysLysAlaProGly...AlaSerProGlnVasProLeuSerP 1358

4631 TTGGAGCAAGAGCGCCCTCACACCGGACAGGAGAGCTGCTGC 4680
::: ::: |||::: |||||::: |||
1358 roserleuserasplysasparglyleuylsglyalaglyserpro 1374
|||::: ::: ::: ||| ||| |||::: |||||
4681 CGGCTCGAGCGCCCTCGCGCGGAGGAGCCAGCGGCTGTCACCG 4730
1375 valglyvalglyleuylleuvalasnvalglyhrcglylnysleupr 1391
::: :::
4731 GCG.....AACGGGC 4741
1391 cthrserylalaaeproleucysarasnprothrarnrserleu 1408
||| ||| ||| |||::: |||||
4742 AGGCGCGGAGCGAGCGCTTCTGTCACCGGCAAGGAGCGAGCGGT 4791
1408 ysglylysleuethasrserlylsleuser..... 1418
::: :::
4792 CGCCATGGGCGAGAACTGCGCGCGCGCGCTGTGCGCGCGCC 4841
1419 serthraspcysphelythrnglualeuethrser. ProgluAlaLeu 1435
||| ||| |||::: ||||| |||
4842 TCGACACCGGTACGCGCGCTCGACCGTCACTCGACCG..... 4881
1435 InproglyglythrAlaLeuAlaProlylsArgserArglyArg 1451
||| ||| ||| |||::: |||||
4882GCCGCTGGGAGATCGTCCGCGGAGGA 4914
1452 AlaGlyAlaHis.....G1 1456
||| ||| ||| |||::: |||||
4915 GCTGACCTCACCGCGGTACACCAAGCCCGCTTCTGCGAGTGG 4964
1456 ylsuserlysglyproleuylulysArprotyrleuylProAla..L 1472
||| ||| ||| |||::: |||||
4965 CGCTGTCGCGCTCTCTCGACACACAGCGCTCGCCGACGTGTCACC 5014
1472 euleuLeuThrProArAsparg..... 1479
||| ||| ||| |||::: |||||
5015 GGGCACTCGGTGGGAGATCGCGCGCGGACGTCGCGGTGCTCTC 5064
1480Alase 1481
5065 COTGACGAGCGCGACGTCGTCAACCGCGCGCGGTCACTCACT 5114
1481 rglThrngluAlaasergluasasnserglvglyglylylsyP 1498
||| ||| ||| |||::: |||||
5115 CGGCGCGGAGGCGCGCGGTGTCGCGGTGACGCGG..... 5154
1498 rplysmetgluLeuLeuylProAlaserInProProgluylArgPro 1514
||| ||| ||| |||::: |||||
5155CGAGCGGAGGTGTCGAGTCCGTGAAGGCGTACGA 5190
1515 CysglInProglInThrArgAlaGlnLysGlnProglYHsrhnsenYrse 1531
||| ||| ||| |||::: |||||
5191 GGGGAGGGTGCGGTGCGCGCGGTCAACG.....ACCGACGCGCG 5231
1531 rserTysSerlyArglysargleuThrArglyAlaIasasnThrT 1548
::: ||| ||| ||| |||::: |||||
5232 TGGTGTCTCGGCAAGCGGAGCGCGCGGAGATCCGCGGTATGG 5281
1548 hrSerSer..... 1550
::: :::
5282 GCGGAGACGCGCGCGGACACCGGAGGCTGCGGTACGACGAGCTTCA 5331
1551 ..ProCyslyselYArgAlaLysArgArgGlnGlnInValleupr 1566
||| ||| ||| |||::: |||||
5332 CTCCCGCACATGAGAGAGCTCTCGACGA.....GTCC 5366
1566 oleuAspProAlaGlnProglulle. ArgLeuYsTyrIlesericy 1582
||| ||| ||| |||::: |||||
5367 TCGGTCGCCGAGAGGCTGACCTCGAGGAGCGGAGATCCCGTGTGC 5416
1582 slysrArgLeuArgSerAspSerArgThrProAlaPheSerProPheVala 1599
||| ||| ||| |||::: |||||

5417 TCACGGTCAACGCGCGCTGTACGTGCGGAGCTCACTTCCGCCG 5466
1599 rglValGlnLysArgAspAlaPheThrIleCysThrValIasnsr 1615
||| ||| ||| |||::: |||||
5467 GT.....ACTGGGTCAACAGATCCGCGGC 5492
1616 Pro.....GlyAspAlaProLysProHisArg..LysPr 1626
||| ||| ||| |||::: |||||
5493 CGGTGCGGTCTCGAGCGCGGTGCGACACCTGGCGCGCCAGAGCGAGC 5542
1626 oserSerSerIasSerSerSerSerSerSerPheSerLeuAspa 1643
||| ||| ||| |||::: |||||
5543 GTCTGTCGAGATCGCGCGCGGTCTCTACGCGCATCGCGGAGGA 5592
1643 lalaglyAlaSerleuAlaThr.....LeuProglyIysr..... 1655
5593 GGCCTGCGCGCGGAGCGAGCGCGCGGAGCGGAGCTGCGGTG 5642
1656IleLeuGlnProArgProSerLeuProle 1665
5643 TCCCGCTCTGCGCGCGCGCGCGCGCGGAGCGCGCGCGGT 5692
1665 user.....SerThmetHis,LeuGlyProValaSerlys 1677
||| ||| ||| |||::: |||||
5693 CTCGCGACCGCGCATGTCCACGCGCGCGCTTGGACCGGCGTC..... 5736
1678 AlaLeuSerThrSerCysleuValCysCysleuCysGlnAsnProAlas 1694
||| ||| ||| |||::: |||||
5737GTCTTCCGAGAGCGCGCGCGCGGAGCTGCGCA 5771
1694 nPheLys.....AspLeuGlyAspLeuLysCysglyProtyr. 1705
||| ||| ||| |||::: |||||
5772 CGTAGCGCTTCCGCGCGGAGCACTACTGCGTGAACGCGCGCGGTACG 5821
1706 ..TyrProglInHisCysleu..... 1711
||| ||| ||| |||::: |||||
5822 GAGCGCGCGCTGCTGACCGCGCGCGGACCGCGCTGACAGC 5871
1712ProlylsLysProLysle 1718
5872 CACGTGAGGTCCCGCGCGGAGAGCGGTCTGCTGACGCGCGGTCT 5921
1718 ulysglulysVal.....ArgProgluGlyThrCysgluAla 1731
||| ||| ||| |||::: |||||
5922 CCCTGACGAGACACCGCTGCGCGGCGGACGACAGGT..... 5958
1731 laseleuProleuGlnArgThrLeuLysglYProglucysAlaAla 1747
||| ||| ||| |||::: |||||
5959CAAGGCGCGGTCTGCTGCGCGGCGCGCGCTTCTGAGCT 6000
1748 AlaThrAlaGlyLysProProArg 1755
||| ||| ||| |||::: |||||
6001 CGCCTCGCGCGGCGGAGCACCGT 6024

Percent Similarity: 87.755 Percent Identity: 75.510

alignment_block:

US-09-508-821B-6 x A1893920

Align seg 1/1 to: A1893920 from: 1 to: 963

21 SerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnProSerGlnAlaG1 37
 1 TCCAGAGAGATCGCCCTGGAGAACTACAGCAGCCGGGTCAGGCTGG 50
 37 yLeuSerCysAspArgGlnArgLeuAlaLysAspTyrTyrAsnProG 54
 51 GCTACGCTGATCGCAGCGGCTGCTGCCAGGACTACTACAGCCCTC 100
 54 lnProTyrProSerTyrGlnGlyAlaGlyThrProSerGlyThrAla 70
 101 AGCCCTATACAGGCTATGAGGTGGCACTGTACACTTCTGGCAGGTG 150
 71 AlaAlaValAlaAlaAspLysTyrHisArgGlySerLysAlaLeuPro 87
 151 GCCACAGCAGCTGACGACAACTGACACGAGGAGCAGCAATCC..... 192
 87 rGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTyrGlyValGln 104
 193CTGCAGGGAGGCGGAGCTTCCCGAGCTAT...GTCAAG 229
 104 spSerSerProTyrProGlyArgTyrAlaGlyGlnGlySerLeuGlnAla 120
 230 ACAGCAGCCCTACCCAGGGGCGTACTCTGGGGAGGAGGAGTCTTACAGC 279
 121 TrpGlyAlaProGlnProProProProGlnProGlnProLeuProAlaG1 137
 280 TGGGGGGGGCCACAGCCAGCCCTCTCAGCCACAGCCTCTGCCCGGGGC 329
 137 yAlaAlaLysTyrAspGlnLeuMetLysTyrThrAlaVal..... 153
 330 AGTGACGAGTATGAGGAGAACCTGTATGAGGAGGAGCAGCTTGTGCTCCTC 379
 153 rSerArgGlnTyrAlaGlnGlnGlyAlaGlnValProPheArgThrHis 169
 380 CAAACAGGCACTACCTGAGCAGGCGCCCGCCAGCTTCCCTTTCGAGCTAC 429
 170 SerLeuHisValGlnGlnProProProProProGlnProGlnProLeuAlaTyrPr 186
 430 AGCTTGCATGTC...CCACCACACAGCCTCAGCAGCCCTGGCTTACCC 476
 186 oLysLeuGlnArgGlnLysLeuGlnAsnAspLysLeuSerProLeuProp 203
 477 CAACTCCAAAGCAGAAACACACAGCAGCCTTGTGCTCCTGCTGCCCT 526
 203 heProGlnGlyThrHisPheProGlnHisSerGlnSerPheProThrSer 219
 527 TCCCGAG..GGCAGCCACTTCCCGCAGATTCACGATTCCTCCCTACCTCC 575
 220 SerThrTyrSerSerSerValGlnGlyGlyGlnGlyAlaHisSerTy 236
 576 TCCACTTATGCCCAAGAGTGGGCTGGTGGGCGGCGGCGGCTCCCT 625
 236 rLysSerCysThrAlaProThrAlaGlnProHisAspArgProLeuThrA 253
 626 ACAGAGCTGCACACACACACTGCGCCAGCTCATGATAGCGCGATGAGTG 675
 253 laSerSerSerLeuAlaProGlyGlnArgValGlnAsnLeuHisAlaTyr 269
 676 CCAATGCGAACCTGGCTCAGGCGNACGGGCTCCAGAAATCTCACCTTAC 725
 270 GlnSerGlyArgLeuSerTyrAspGlnGlnGlnGlnGlnGlnGlnGln 286
 726 GCGCTGGCGGCTTGGCTAC..... 746
 286 nGlnGlnGlnGlnAlaLeuGlnSerArgHisHisAla..GlnGlnThr 302

747..GAGCAGCAGCAGCAGCAGCAGCTTCAAGCGCGTCAACACACACCCAGGAACA 795
 803 LeuHisTyrGln.AsnLeuAlaLysTyrGlnHisTyrGlyGlnGlnGlyG 319
 796 CTTCACTACAGAACTGGCCAGTACCACTATGGACACAGGATAC 845
 319 lnglyTyrCysGlnProAspAlaAlaValArgThrProGlnGlnTyr 335
 846 AGGCTACTGTCCACCGGACACAGCTGTCCAGACTCCACACAGGATAC 895
 336 GlnThrPheSerProSerSerHisSerProAlaArgSerValGly.A 352
 896 GAGACTTTCAGCCGCGACTCCAGCCATCCCTGACGT...CTGGGTTC 942
 352 rgSerProSerTyrSerSerThr 359
 943 GT...CCCTTCTATAGCTCCACC 962

seq_name: gb_est2:BE869510

seq_documentation_block: 865 bp mRNA EST 20-OCT-2000
 LOCUS BE869510
 DEFINITION 601446178F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850146 5',
 mRNA sequence.

ACCESSION BE869510
 VERSION BE869510.1 GI:10318286
 KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 865)

AUTHORS NTH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgaabs-re@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM9568 row: 1 column: 19

High quality sequence stop: 662.

Location/Qualifiers

source

1..865
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3850146"
 /clone_lib="NIH_MGC_65"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: NotLi
 site:2: SalI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

BASE COUNT 184 a 294 c 242 g 145 t
 ORIGIN

alignment_scores:

Quality: 1088.50 Length: 260
 Ratio: 4.612 Gaps: 4
 Percent Similarity: 90.769 Percent Identity: 85.385

alignment_block:

US-09-508-821B-6 x BE869510

Align seg 1/1 to: BE869510 from: 1 to: 865

1499 LysMetGlnGlnLeuGly...ProAlaSerGlnProProGlnGlyArgPr 1514

```

|||||
1 AAGATGAGAGAGCTGGGCTGCTCCCTCCAGCCCGCCGCGGAGGCGCAGGCC 50
1514 OCYGLINPROGINTHRATGALAGLNLGSLNPROGLYHSTRHSPNTYS 1531
|||||
51 CTGGCAGCCCGCAGACAGGCGACAAAGACCCAGCCGACACCACTACCA 100
1531 eSerTyrSerLysArgLysArgLysArgLysArgLysArgLysArgLys 1547
|||||
101 GCAGCTATTCACAGCGAGAGCCCTCACTCGGCGCGGCGCCAGAACACAC 150
1548 ThrSerSerProCysLysGlyArgLysArgLysArgLysArgLysArgLys 1564
|||||
151 ACCTTTCACCTGCTGAGGCGCTGAGGCGCGAGGCGAGGCGAGGCGAGG 200
1564 LLeuProLeuAspProAlaGluProGluLLeuArgLysTyrLLeuSer 1581
201 GCTGCCCCCTGGATCCCGAGAGCCCTGAATCCCGCTCAAGATACATTCCT 250
1581 eTCysLysArgLysArgLysArgLysArgLysArgLysArgLysArgLys 1597
|||||
251 CTGGCAGGCGGCTGAGGCTGAGGCGGAGGCGGAGGCGGCTTCTCAGCTTC 300
1598 ValArgValGluLysArgAspAlaPheThrLLeuCysThrValValAs 1614
301 GTGCGGAGTGGAGAGCGAGCGAGCGCTTCACGACATGACAGCTGTTCGA 350
1614 nSerProGlyAspAlaProLysProHisArgLysProSerSerSerLAs 1631
351 CTCCTCCGAGATGCGCCCAAGCCCGCAGGAGGAGGAGGAGGAGGAGGAG 400
1631 eSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1647
401 CTTCTCCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 450
1648 LeuAlaThrLeuProGlyLysLLeuGlnProAlaGProSerLeuSer 1664
451 CTGGCCAGAGTCCCTGGAGGCTCCATCTCGAGGCTCGGCTCGCTTGGCC 500
1664 OLeuSerSerThrmethIleuGlyProValValSerLysLysLLeuSer 1681
501 CTTCTCTCCAGATGACATGAGGCGCTGAGGCGCTTCCAGAGGCGCTGAG 550
1681 hSerCysLysLeuValCysCysLysCysGlnAsnProAlaAsnPhelys 1697
551 CTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
1697 PLeuGlyAspLeuCysGlyProTyrTyrProGluHisCysLeuProLys 1714
601 CTTGCTGAGCTCTGCTGAGGCTCTACTACTGAGACACTGACTGCGCACA 650
1714 yLysProLysLeuLysGluLysValArgProGluGlyThrCysGluGlu 1730
651 AGAAGCAAACTCCAGAGAGAGTGGCGGCGAGAGGCGACCTGGAGAGGA 700
1731 AlaSerLeuProLeuGluArgThrLeuLysGlyProGluLysAlaAla 1747
701 GCTCTGCTGCGCTGAGAGAGACATCAAAAGTCTCCGAGTGTGGAAGTTC 750
1747 AlaAlaThrAlaGlyLysProPro 1754
751 CCGAGACGTGGCGGAGAGGCCCA 774
seq_name: gp_est2:B6828494
seq_documentation_block:
LOCUS B6828494 724 bp mRNA EST 22-MAY-2001
DEFINITION 602751486F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4904229 5',
mRNA sequence.
ACCESSION B6828494
VERSION B6828494.1 GI:14176081
KEYWORDS EST.
SOURCE human.

```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.fremail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMIL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMIL at:
http://image.llnl.gov
Plate: LINC1803 row: 1 column: 22
High quality sequence stop: 724.
FEATURES
source
location/Qualifiers
1..724
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4904229"
/clone_id="NIH_MGC_17"
/tissue_type="chondrosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOT8; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies). "
BASE COUNT 161 a 246 c 217 g 100 t
ORIGIN
alignment_scores:
quality: 996.50 length: 211
ratio: 4.982 gaps: 5
Percent Similarity: 94.787 Percent Identity: 93.365
alignment_block:
US-09-508-821b-6 x B6828494 ..
Align seg 1/1 to: B6828494 from: 1 to: 724
1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnAsnTyr 17
|||||
101 ATGCAAGTCTTTTCGAAAAGGTGGTTCACAGGCAAAACACAGACTA 150
17 rgInGlnThrSerGlnGluThrSerArgLeuGluAsnTyrArgGln. Pr 33
|||||
151 CCGAGAGCTCCGAGAGAACATCACTCCATGAGATTCAGTGGACTA 200
33 oSerGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspT 50
201 GAGTCAGGCGCGGCTAAGCTGAGACCGGCGGCTGCTGCCAGAGACT 250
50 yTyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyThrPro 66
251 ATTATTAACCGCAGCCTTACCGGAGCTATGAGGAGGCGGCGGACAGCC 300
67 SerGlyThrAlaAlaValAlaAlaAspLysTyrHisArgGlySerLys 83
301 TCGGCACTGAGAGCGCGGCTGCGCGGAGACATCAACAGCAGAGCA 350
83 sAlaLeuProThrGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyT 100
351 GGCCCTGGCCCAAGAGAGGCTGCGAGGAGGAGGCGGCTTCCCGGCT 400
100 yGlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGln 116

```

```

|||||
401 ACGGGTCCAGGACAGAGCCCTTACCCAGCGGTATGCTGGTGAGAG 450
|||||
117 SerLeuGlnAlaTrpGlyAlaProGlnProProProGlnProGlnPr 133
|||||
451 AGCCTTCAGGCTTGGGGGCCCCCAGCAGCCACCAACCCACAGCCGAGCC 500
|||||
133 oLeuProAlaGlyValAlaLysTyrAspGlnAsnLeuMetLysThrA 150
|||||
501 ACTACTGCAGGGGTGGCAAGTATGATGAGAACTTGATGAAAAAGACAG 550
|||||
150 laValProProSerArgGlnTyrAlaGlnGlnAlaGlnValProPhe 166
|||||
551 CAGTCCCCCCCCACAGCAGCAGTATCAGACAGAGGGGCCGCCAGTGCCCTT 600
|||||
167 ArgThrHisSerLeuHisValGlnGlnProProProGlnGlnPro 182
|||||
601 CGGACTCACTCCCTTGCACCGTCCAGCAGACACCCGCCGCCAGCAGACC 650
|||||
183 LeuAlaTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAla 199
|||||
651 TGG...CATACCCCAAGCTCCAAAGGAGAGAGTGCAGAAAGACATTGGCT 697
|||||
199 exProLeuProPheProGlnGlyThr 207
|||||
698 CCCCCTGAGCTTCCGCCAGGGGTACC 723

```

seq_name: gb_est2:BF307287

seq_documentation_block:
LOCUS BF307287 1012 bp mRNA EST 21-NOV-2000
DEFINITION 601891563f1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4137399 5',
mRNA sequence.
ACCESSION BF307287
VERSION BF307287.1 GI:11254402
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: AFCC

cDNA library Preparation: Ling Hong/Rubin Laboratory
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCMI044 row: m column: 16
High quality sequence stop: 736.

FEATURES

Location/Qualifiers
1..1012
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4137399"
/clone_lib="NIH_MGC_17"
/tissue_type="inabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pORF7; Site:1: EcoRI;
Site:2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the Laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
260 a 307 c 295 g 148 t 2 others

BASE COUNT
ORIGIN

alignment_scores:
Quality: 989.50 Length: 292
Ratio: 4.039 Gaps: 10
Percent Similarity: 83.904 Percent Identity: 74.658

alignment_block:
US-09-508-821B-6 x BF307287 ..
Align seg 1/1 to: BF307287 from: 1 to: 1012

```

1 MetGlnSerPheArgGlnArgCysGlyPheHisGlyLysGlnGlnAsnTyr 17
|||||
102 AFGCAGTCTTTTCGAGAAAGGTGTGTTTCCATGGCAACACAGAACTA 151
|||||
17 rGlnGlnThrSerGlnGlnThrSerArgLeuGlnAsnTyrArgGlnPro 34
|||||
152 CCAGCAGACCTCGCAGGAAACATCAGCCTAGAGAAATTACAGCAGCCGA 201
|||||
34 erGlnAlaGlyLeuSerCysAspArgGlnArgLeuAlaLysAspTyr 50
|||||
202 GTCAAGGCGGGCTTAAGCTCGACCGCGCGCTGCTGCCAAGGACTAT 251
|||||
51 TyrAsnProGlnProTyrProSerTyrGlnGlyAlaGlyThrProSe 67
|||||
252 TATAACCGCGAGCCTTACCGAGCTATGAGGGTGGCGCTGGCAGCCCTC 301
|||||
67 rGlyThrAlaAlaValAlaLysTyrHisArgLysSerLysA 84
|||||
302 TGGCAGCTGCAGCGCGGTGGCGCGGACAAAGTACCACCGAGCGCAGAG 351
|||||
84 laLeuProThrGlnGlnGlyLeuGlnArgProAlaPheProGlyTyr 100
|||||
352 CCCTGCCGCACACAGAGCGCTGCAGAGGGCGCGCTTCCCTGGCTAC 401
|||||
101 GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGluSe 117
|||||
402 GCGGTCCAGCAGCAGCAGCCCTACCCAGCGCGCTATGCTGGTGAGGAG 451
|||||
117 rLeuGlnAlaTrpGlyAlaProGlnProProProGlnProGlnProL 134
|||||
452 CTTTCAGGCTGGGGGGGCCCCACAGNCAACACCCACAGCGCCAGCA. 500
|||||
134 euProAlaGlyValAlaLysTyrAspGlnAsnLeuMetLysLysThrAla 150
|||||
501 TACCTGAGCGGGTGGCCAGATATGTGAACTTGTGAAAGACAGCA 550
|||||
151 ValProProSerArgGlnTyrAlaGlnGlnGlyAlaGlnValProPhe 167
|||||
551 GTGCCCCCAGC. AGGCAGTATGCAGAGCAGGGCGCCAGGTTCCTTCG 599
|||||
167 qThrHisSerLeuHisValGlnGlnProProProGlnGlnProLeuA 184
|||||
600 GACTCATCCTCCCTCAGCGTCCAGCAGCAACAGCGCGCCAGCGCCCTGT 649
|||||
184 laTyrProLysLeuGlnArgGlnLysLeuGlnAsnAspIleAlaSerPr 200
|||||
650 GATTAACCAAGCTCCAAAGCAGAGAGCTTGCAGACAGACATNGSCTCCC 699
|||||
200 oLeuProPheProGlnGlyThrHisPheProGlnHisSerGlnSerPhe 216
|||||
700 TCTTGGCTTCCCA...GGGTACCACTTTCCTCAGGATTCAGTCTT... 743
|||||
217 ProThrSerSerThrTyrSerSerValGlnGlyGlyGlnGlyAl 233
|||||
744 ...TCCAGACTCTTCCATATCTCTCTCTGTCAGTGGGGGAGCGGG... 788
|||||
233 ahisSerTyrLysSerCysThrAlaProThrAlaGlnProHisAspArgp 250
|||||
789 .CGCATCTATAAGAGT.....TGACAGACGAATGGCAAC 822
|||||
250 roLeuThrAlaSerSerSerLeu...AlaProGlyGlnArgValGlnAsn 265
|||||

```


Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et3-RC0-CT0379-290)
100-032-clit3-2000-01-29at4-1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 566.
Location/Qualifiers

FEATURES

1. 566
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0379"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 99 a 211 c 149 g 107 t

ORIGIN

alignment_scores:

Quality: 819.00 Length: 160
Ratio: 5.184 Gaps: 1
Percent Similarity: 98.750 Percent Identity: 98.750

alignment_block:

US-09-508-821b-6 x AW860154

Align seg 1/1 to: AW860154 from: 1 to: 566

1597 PheValArgValGluLysArgAspAlaPheThrThrIleCysThrValva 1613
|||||
13 TTCGTGGGGTGGGAGGAGGAGCGGTTCACCCACCATATGACACTGTGT 62
|||||
1613 LasSer.ProGlyAspAlaProLysProHisArgLysProSerSer 1629
|||||
63 CAATCCCGTGGAGATCGCCCAAGCCAGCCAGAGCGCTTCTCTCTCT 112
|||||
1630 AlaSerSerSerSerSerSerPheSerLeuAspAlaIylAl 1646
|||||
113 GCCT 162
|||||
1646 aSerLeuAlaThrLeuProGlyGlySerIleLeuGlnProArgProSerL 1663
|||||
163 CTCCCTGGCCACTCCCTGGAGGCTCCATCTGTCGAGCCGGCCCTCT 212
|||||
1663 enProLeuSerSerThrMethisLeuGlyProValValSerLysAlaLeu 1679
|||||
213 TCCCTCTCTCTCCAGATGACATGGGGCTGTGTGTTCAGAGCCCTG 262
|||||
1680 SerThrSerCysLeuValCysCysLeuGlnAsnProAlaAsnPhely 1696
|||||
263 AGTACTCTTGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 312
|||||
1696 sAspLeuGlyAspLeuGlyProGlyProGlyProGlyProGlyProGly 1713
|||||
313 GGACCTTGGGACCTCTGTGGCCCTACTACCTTGACACTGCTCTCCCA 362
|||||
1713 yslLysLysProLysLeuLysGluLysValArgProGluGlyThrCysGlu 1729
|||||
363 AAAAGAGCCAAACCTCAAGGAGAGGTGGGCCAGAGGACCTGTGAG 412
|||||
1730 GluAlaSerLeuProLeuGluArgThrLeuLysGlyProGluCysAlaAl 1746
|||||

413 GAGGCTCGCTGCCGCTTGAGAGACACTCAAGAGTCCGAGTGTGCAGC 462

1746 aAlaAlaThrAlaGlyLysProProArg 1755

|||||
463 TGCCGCCACTGCCGGAAGCCGCCAGG 490

seq_name: gb_est1:AW860199

seq_documentation_block:

LOCUS AW860199 566 bp mRNA EST 19-MAY-2000
DEFINITION RC0-CT0379-060200-032-cl1 CT0379 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW860199
VERSION AW860199.1 GI:7955892
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1. (bases 1 to 566)

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et3-RC0-CT0379-060)
200-032-clit3-2000-02-06at4-1)
Seq primer: puc 18 forward
High quality sequence start: 11
High quality sequence stop: 566.

FEATURES

source

1. 566
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0379"
/dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

99 a 211 c 149 g 107 t

ORIGIN

alignment_scores:

Quality: 819.00 Length: 160
Ratio: 5.184 Gaps: 1
Percent Similarity: 98.750 Percent Identity: 98.750

alignment_block:

US-09-508-821b-6 x AW860199

Align seg 1/1 to: AW860199 from: 1 to: 566

1597 PheValArgValGluLysArgAspAlaPheThrThrIleCysThrValva 1613
|||||

13 TTGCTGGCGGTGAGAGAGAGAGCGGTTCACACATATGCACTGTGT 62
 1613 Iasnsr.ProgIyaspAlaProIySahgIyPProSerSer 1629
 63 CAACCTCCGCTGAGATGCGCCCAAGCCCAAGAGAGAGCTTCCTCT 112
 1630 AlasrSerSerSerSerSerSerSerSerSerSerSerSerSer 1646
 113 GCCTCCTCTTCCTCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 162
 1646 aserleuAlaThrleuProIyglYserIleuGlnProIyPProSer 1663
 163 CTCCCTGGCCACCTCCCTGAGAGCTTCCTGCTGCTGCTGCTGCT 212
 1663 eupProIySerSerTherMetHisleuGlyProValIySerIyS 1679
 213 TGCCCTCTCTCCACGATGCTGAGGCTGAGGCTTCCAGGCTGCT 262
 1680 SerTherSerCysLeuValCysIySleuGlnsnpProIySahp 1696
 263 AGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 312
 1696 asPleuGlyaspLeuGlyPProIyTyrProIyHIScysLeuPro 1713
 313 GGACCTGGGAGACTCTGTGAGCCCTGCTGCTGCTGCTGCTGCTG 362
 1713 YsIySlyPProIySleuGlySlyValIyPProIyglYThrCysG 1729
 363 AAAAGAGCCAAACTCAAGAGAGAGTGGCGCAGAGAGAGCTGTG 412
 1730 GlnAlaSerleuProIyGlnIyThrIySlyGlyPProIyGlyS 1746
 413 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 462
 1746 aAlaAlaThrAlaGlyIySlyPProIyArg 1755
 463 TCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 490

seq_name: qb_gss:A2642061

seq_documentation_block: 501 bp DNA GSS 14-DEC-2000
 LOCUS A2642061 1M0504124R Mouse 10kb plasmid UNGCJM library Mus musculus genomic
 DEFINITION clone UNGCJM0504124 R, DNA sequence.

ACCESSION A2642061
 VERSION A2642061.1 GI:11768322

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Mouse whole genome scaffolding with paired end reads from 10kb
 plasmid inserts
 Unpublished (2000)
 Contact: Robert B. Weiss
 University of Utah Genome Center
 University of Utah
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SEC. UT
 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu
 Insert length: 10000 Std Error: 0.00
 Plate: 0504 row: 1 column: 24
 Seq primer: CACACAGAGAAAGAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 501.
 Location/Qualifiers

SOURCE

1. 501
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UNGCM0504124"
 /clone_11b="Mouse 10kb plasmid UNGCJM library"
 /sex="Male"
 /lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv. Purified genomic DNA from M.
 musculus C57BL/6J (male) was obtained from the Jackson
 Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA
 was hydrodynamically sheared by repeated passage through a
 0.005 inch orifice at constant velocity. The sheared DNA
 was blunt end-repaired with T4 DNA polymerase and T4
 polynucleotide kinase. Adaptor oligonucleotides were
 ligated to the blunt ends in high molar excess. The
 ligated DNA was purified and size-selected for a 9.5 to
 10.5 kb range using preparative agarose gel
 electrophoresis. Vector DNA was prepared from a derivative
 of PMD42 (g14732114[gb]AF129072.1), a copy-number
 inducible derivative of plasmid R1. The vector was ligated
 with adaptors complementary to the insert adaptors and
 purified. The sheared, adaptor mouse DNA was annealed to
 adaptor vector DNA, and transformed into
 chemically-competent E. coli XL10-Gold (Stratagene) cells
 and selected for ampicillin resistance."

alignment_scores:

Quality: 807.00 Length: 165
 Ratio: 4.981 Gaps: 0
 Percent Similarity: 98.182 Percent Identity: 93.333

alignment_block:

US-09-508-821b-6 x A2642061

Align seq 1/1 to: A2642061 from: 1 to: 501

350 ValGlyATGserProSerTyrSerSerThrProSerProIyMetProAs 366
 3 GTGGGTCGCTCCCTCTTATGCTGCTGCTGCTGCTGCTGCTGCTG 52
 366 nleuGlnsnpPProIySerGlnGlnProIySerThrGlyAlaPhep 383
 53 TCTGGAGACTCCCTATAGCCAGCGCGCTTATAGTACGGGCTTCC 102
 383 roAlaGlyIleThrAspHisSerHisPheMetProIyLeuAsnProSer 399
 103 CCACAGGCAATCAACACAGCGACTTATGCCCGCTTAAACCGCTCC 152
 400 ProThrAspAlaThrSerSerValAspThrGlnAlaGlySncysIySyr 416
 153 CCAACAGATGCTGCGAGCTGTGGAGCCGCCAGCGCGCACTGCAGAG 202
 416 cleuGlnIySlySlySleuProIySleuIySleuSerAspIySerLeuG 433
 203 CTTGCAAAAGAGAGAGCTGCTGCAACTTGTCTGCGACTGAGCTGCT 252
 433 InSerleuThrAlaThrleuGlnIyValGlnsnpIleSerAsnThrVal 449
 253 AGAGCTTCACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302
 450 GlnGlnIySleuIySleuSerIySlyAlaIyValProIyGlnIySlyGlyIy 466
 303 CAGAGCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 352
 466 asPleuValSerArgThrProIyGlnIySlySerGlnIySlySerP 483
 353 GAACCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402

483 rogluGlySerGlyTyrSerAlaGluProAlaGlyThrProLeuSerGlu 499
 |||||
 403 CTGAGGAGCGGCTCTTCAGCTGAGCGAGCGGCACACCGCTCTGAG 452
 |||||
 500 ProProSerSerThrProGlnSerThrHisAlaGluProGlnGlu 514
 |||||
 453 CGCGGAGCAGCAGCAGCAGTCCACCGCTGAGCCACAGGAC 497
 |||||

seq_name: gb_est1:AW860153

seq_documentation_block: 558 bp mRNA EST 19-MAY-2000
 LOCUS AW860153
 DEFINITION RCO-CT0379-290100-032-b08 CT0379 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW860153
 VERSION AW860153.1 GI:7955846
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 558)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RCO-CT0379-290
 100-032-b08t3-2000-01-29t4-1)
 Seq primer: puc 18 forward
 High quality sequence start: 38
 High quality sequence stop: 558.

FEATURES

source Location/Qualifiers

1..558
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CT0379"
 /dev_stage="Adult"

/note="Organ: colon; Vector: puc18; Site 1: SmaI; Site 2:
 SmaI; A mini-library was made by cloning products derived
 from ORSTES PCR (U.S. Letters Patent application No. 196
 ,716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 98 a 207 c 147 g 106 t
 ORIGIN

alignment_scores:
 Quality: 794.00 Length: 161
 Ratio: 4.994 Gaps: 1
 Percent Similarity: 98.758 Percent Identity: 97.516

alignment_block:

US-09-508-821B-6 x AW860153

Align seg 1/1 to: AW860153 from: 1 to: 558

1506 PropheValArgValGluLysArgAspAlaPheThrThrIleCysThrVa 1612
 |||||
 13 CCTCTGCTGGGTGGAGAGGGA. GACGGCTTCCACCAATGACACTGT 51
 |||||
 1612 lValasnSerProGly AspAlaProLysProHisArgLysProSerSer 1628
 |||||
 52 TGTCAACTCCCTGTGAGATGGCCCAAGCCCAAGAGAGCCTTCTCC 101
 |||||
 1629 SerAlaSerSerSerSerSerSerSerPheSerLeuAspAlaIaGl 1645
 |||||
 102 TCTGCTCTCTCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCT 151
 |||||
 1645 yAlaSerLeuAlaThrLeuProGlyGlySerIleLeuGlnProArgProS 1662
 |||||
 152 GGCCTCCCTGGCCACACTCCCTGGAGGCTCCATCTGACGCCGCGCCT 201
 |||||
 1662 erLeuProLeuSerSerThrMetHisLeuGlyProValValSerLysAla 1678
 |||||
 202 CTTGGCCCTCTCTCTCCACGATGACTTGGGGCTGTGTTTCCAAAGCC 251
 |||||
 1679 leuSerThrSerCysLeuValCysCysLeuGlnAsnProAlaAsnPh 1695
 |||||
 252 CTGAGTACTCTTGGCTTGTGTGCTTGTGCTTGTGCTTGTGCTTGT 301
 |||||
 1695 eLysAspLeuGlyAspLeuCysGlyProTyrTyrProGluHisCysLeuP 1712
 |||||
 302 CAAGGACTTGGGGACCTCTGTGCCCCCTACTACCTGAACACTGGCTCC 351
 |||||
 1712 roLysLysLysProLysLeuLysGlyLysValArgProGluGlyThrCys 1728
 |||||
 352 CCAGAGAGAGAGGCAAACTCAAGGAGAGAGTGGCGCCAGAGGACCTGT 401
 |||||
 1729 GluGluAlaSerLeuProLeuGluArgThrLeuLysGlyProGluCysAl 1745
 |||||
 402 GAGGAGGCTCGCTGGCTGGCTTGTGAGAACACTCAAGGTCCTCCGAGTGTGC 451
 |||||
 1745 aAlaAlaThrAlaGlyLysProProArg 1755
 |||||
 452 AGCTGCCCCACTGCCGGAAGGCCCCCGAGG 482

seq_name: gb_est1:AW862383

seq_documentation_block: 558 bp mRNA EST 19-MAY-2000
 LOCUS AW862383
 DEFINITION RCO-CT0379-060200-032-b08 CT0379 Homo sapiens cDNA, mRNA sequence.
 ACCESSION AW862383
 VERSION AW862383.1 GI:79558081
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 558)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-RCO-CT0379-060
200-032-b08et3-2000-02-06et4-1)
Seq primer: puc 18 forward
High quality sequence start: 38
High quality sequence stop: 558
Location/Qualifiers

FEATURES

source
1. 558
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0379"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: Sma1; Site_2:
Sma1; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 98 a 207 c 147 g 106 t
ORIGIN

alignment_scores:
Quality: 794.00 Length: 161
Ratio: 4.994 Gaps: 1
Percent Similarity: 98.758 Percent Identity: 97.516

alignment_block:
US-09-508-821B-6 x AW862383 ..

Align seg 1/1 to: AW862383 from: 1 to: 558

1596 prophetaalargyalgluysargaspalaphethrthilicyslhyra 1612
|||||
3 CCTTCGTGGGGGTGGAGAGGA. GACGCTTACCAACATATGCACTG 51

1612 ValanserProGly. AspalaprolsProhlsArglyProSerSer 1628
|||||
52 TGTCAACTCCCTCTGTGATGCGCCCAAGCCCAAGAGGCTTCTCC 101

1629 SerAlaserSerSerSerSerSerSerSerSerSerSerSerSer 1645
|||||
102 TCGGCTCCTCTCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 151

1645 yAlaserleuAlathleuProGlySerleleuglinProArpPro 1662
|||||
152 GGCCTCCCTGGCCACACTCCCTGGAGGCTCCATCTGCGAGCCGCC 201

1662 etleuProleuSerSerThMetHlsleuglyProValSerlyAla 1678
|||||
202 CCTTGCCCTCTCTCTCCAGATGCACTTGGGCGCTGGTTCCAGGCC 251

1679 leuSerThSerCysleuValCysCysleucysGlnAsnProAlasRph 1695
|||||
252 CTGAGTACCTCTGCTGCTTCTCTCTCTCTCTCTCTCTCTCTCT 301

1695 elysaspieuglyaspieucysglyProTyrrTyrrProGluHlsCysleuP 1712
|||||
302 CAAGGACCTTGGGGACCTCTGTGGCCCTACCTACCTGAGACCTGCTCC 351

1712 rolyslslySProlyslsleuysGluysValarProGluGlyThrcys 1728
|||||
352 CCAAAAGAGAGCCAAACTCAAGAGAAAGAGAGAGAGAGAGAGAGAG 401

1729 GluGluAlaserleuProleugluArgthleuysGlyProGluCysAl 1745
|||||
402 GAGGAGGCTCCGCTGCGCTGAGAGAAACACCAAGAGTCCGAGTGC 451

1745 aAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1761
|||||
452 AGCTGCGCCGACCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 482

seq_name: gb_est1:AA052617

seq_documentation_block:
LOCUS AA052617 611 bp mRNA EST 13-SEP-1996
DEFINITION mc82a07.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA
clone IMAGE:355284 5' similar to gb:029801 Mouse mRNA for unknown
product, complete cds (MOUSE);, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
JOURNAL
COMMENT
The Nasu-HHMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through INMIL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:227084

Trace considered overall poor quality
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 1.
Location/Qualifiers

FEATURES

source

1. 611
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:355284"
/clone_lib="Soares mouse embryo NM013.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGAGGAGGCGGCGGAGAAATTTTCTTTTCTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT733 vector. Library went through one round of
normalization, and was constructed by Benito Soares and
M.Fatima Bonaldo."

BASE COUNT 139 a 204 c 165 g 103 t
ORIGIN

alignment_scores:
Quality: 793.50 Length: 219
Ratio: 4.312 Gaps: 7
Percent Similarity: 84.018 Percent Identity: 72.146

alignment_block:
US-09-508-821B-6 x AA052617 ..

Align seg 1/1 to: AA052617 from: 1 to: 611

22 GluGluThrSerArgleuGluAsnTyArgGlnProSerGlnAlaGlyLe 38
|||||
|||||

2 CAGGAGACATCGCGCTGGAGAACTACAGCAGCGCGGTACAGCGTGGGCT 51
 38 uSerCysAspArgGlnAlaLeuAlaLysAspTyrTrsProGlnP 55
 52 AGCTGTGTATCGGACGCGCTGCTGGCCAGGAGCTACTACAGCCCTCAGC 101
 55 rOTyProSerTyrGluGlyAlaGlyThrProSerGlyThrAlaAla 71
 102 CTTATACAGGCTATGAGGTGGCTGTGTACAGCTTCTGGCAGCGTGGCC 151
 72 AlaValAlaAlaAspLysTyrHisArgLysSerLysAlaLeuProThrG 88
 152 ACAGCAGGTGAGACAGTACACCGAGGAGGAGATCC..... 190
 88 nGlnGlyLeuGlnGlyValProAlaPheProGlyTyrGlyValGlnAsp 105
 191CTGCAGGGGAGGCGAGCTTCCCACTAT.....GTCAAGACA 230
 105 eSerProTyrProGlyValTyrAlaGlyGluGlnSerLeuGlnAlaTrp 121
 231 GCAGCCCTACATAGGCGCTACTCTGCGAGGAGAGTCTTCAGACCTGG 280
 122 GlyAlaProGlnProProProGlnProGlnProGlnProAlaGlyVa 138
 281 GGGGGCCACAGACCCCGCTCTCAGCCACAGCTCTGCCGGGGCAGT 330
 138 LAlaLysTyrAspGluAsnLeuMetLysLysThrAlaValPro...ProS 154
 331 GAGCAGATGAGGAGGAGCTGTGATGAGNAGCAGTGTGCTCTTCCAA 380
 154 eArgGlnTyrAlaGluGlnGlyAlaGlnValProPheArgThrHisSer 170
 381 ACAGGCACTACCTGAGCAGGCGCCCGAGCTTCCGCTTCCGAGCTCACAGC 430
 171 LeuHisValGlnGlnProProProGlnGln...ProLeuAlaTyrPro 186
 431 CTGCATGC.....CCACACACAGCTCAGACCCCTGGCTTACGCA 474
 187 LysLeuGlnArgGlnLysLeuGlnAsnAspLysAlaSerProLeuProPh 203
 475 AAATTCNAGGCGAARACAGAGAGGAGCTTCCGCTTCCGCTTCCGCTT 524
 203 eProGlnGlyThrHisPheProGlnHisSerGlnSerPhe..... 216
 525 CGCAGGAGGAGCACTTTCAGCAGCTTCCGCTTCCGCTTCCGCTTCC 574
 217ProThrSerSerThrTyrSerSerValGlnGlyGly 230
 575 ACTATGCCCCAATGTTCA.....GGTGGTGGC 603
 231 GlnGly 232
 604 AAGGG 609

seq_name: gb_est2:BF304773

seq_documentation_block: 832 bp mRNA EST 21-NOV-2000
 LOCUS BF304773 832 bp mRNA EST
 DEFINITION 601888356f1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4122469 5',
 mRNA sequence.

ACCESSION BF304773
 VERSION BF304773.1 GI:11251671
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 832)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 COMMENT Email: cgabbs-remail.nih.gov

Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLC1005 row: o column: 14
 High quality sequence stop: 619.
 Location/Qualifiers
 1. .832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4122469"
 /clone_lib="NIH_MGC_17"
 /tissue_type="rhodomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pOT7; Site_1: EcoRI;
 Site_2: XhoI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(S). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 195 a 256 c 247 g 134 t
 ORIGIN

alignment_scores:
 Quality: 791.00 Length: 253
 Ratio: 3.821 Gaps: 12
 Percent Similarity: 81.818 Percent Identity: 73.123

alignment_block:
 US-09-508-821B-6 x BF304773 ..
 Align seg 1/1 to: BF304773 from: 1 to: 832

1 MetGlnSerPheArgGluArgCysGlyPheHisGlyLysGlnAsnTyr 17
 101 ATGCGAGTCTTTTCGAGAAAGGTGGTTCCTCATGGCAACACAGAACTA 150
 17 rGlnGlnThrSerGlnGlnThrSerArgLeuGluAsnTyrArgGlnProS 34
 151 CCAGCAGACCTCGCAGGAACATCACGCTAGAGAAATTACAGGCGCGCA 200
 34 eGlnAlaGlyLeuSerCysAspArgGlnGlnArgLeuAlaLysAspTyr 50
 201 GTCAGCGCGGCTAAGCTCGACCGCGAGCGCTGCTCGCCAGGAGCTAT 250
 51 TyrAsnProGlnProTyrProSerTyrGlnGlyValAlaGlyThrProSe 67
 251 TATACCGCGAGCCTTACCGAGCTATGAGGTGGCGCTGGCAGCCCTC 300
 67 rGlyThrAlaAlaValAlaAlaAspLysTyrHisArgGlySerLysA 84
 301 TGGCAGCTGAGCGCGGTGGCGCGACAAAGTACACCGAGGCGAGGAG 350
 84 laLeuProThrGlnGlnGlyLeuGlnGlyArgProAlaPheProGlyTy 100
 351 CCCTGCCACACAGCAAGGCTCGAGGGAGGCGCTTCCCTGGGCTA 400
 100 r.GlyValGlnAspSerSerProTyrProGlyArgTyrAlaGlyGluGlu 116
 401 CTGCGCTCAGGACAGACGCCCTTACCCAGT.GGTATGCTGGTGAGGAG 449
 117 SerLeuGlnAlaTrpGlyAlaProGlnProProProGlnProGlnPr 133
 450 AGCTTCAGGCTT...GGGGGCCCAAGCCCAACCCCATAGCGCAGCAC 496
 133 oLeuProAlaGlyValAlaLysTyrAspGluAsnLeuMetLysLysThra 150
 497 TACCTGCGAGGGGT...GGCAAGTATGATGAGAACTTGTGAAAAAGACAG 543

sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st5-RC0-ST0379-210
100-032-d02at3-2000-01-21at4-1)
Seq primer: puc 18 forward
High quality sequence start: 20
High quality sequence stop: 540.

FEATURES
source
1..540
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ST0379"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site1: Smal;
Site2: Smal; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 97 a 203 c 140 g 100 t
ORIGIN

alignment_scores:
Quality: 785.00 Length: 149
Ratio: 5.268 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.329

alignment_block:
US-09-508-821B-6 x AW821899 ..
Align seg 1/1 to: AW821899 from: 1 to: 540

1607 ThrThrLeuGlyThrValValAlaSerSerProGlyAspAlaProLysProHi 1623
|||||
18 ACCACCATGACACTGTGTCACCTCCCTCGGAGATGGCCCAAGCCCA 67
|||||
1623 sArgLysProSerSerSerAlaSerSerSerSerSerSerPheS 1640
|||||
68 CAGGAGCGCTTCCTCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 117
|||||
1640 erLeuAspAlaAlaGlyValaSerLeuAlaThrLeuProGlyGlySerLe 1656
|||||
118 CCTTGATCGACGCGGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 167
|||||
1657 LeuGlnProArgProSerLeuProLeuSerSerThrMetHisLeuGlyPr 1673
|||||
168 CTGACGCGCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 217
|||||
1673 oValValSerLysAlaLeuSerThrSerCysLeuValCysCysLeuCysG 1690
|||||
218 TGTGCTTCTCAAGCGCTGAGTACCTCTCTCTCTCTCTCTCTCTCTCTCT 267
|||||
1690 InAsnProAlaAspPheLysAspLeuGlyAspLeuCysGlyProTyrTyr 1706
|||||
268 AAAACCGGCCCACTCAAGACCTTGGGGACCTCTGTGGCCCTACTATAC 317
|||||
1707 ProGluHisCysLeuProLysLysLysProLysLeuLysGluLysValar 1723
|||||
318 CTTTACATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 367

Page 13

410 GlnAlaIagIaAsnCySerProLeuGlnIAspIysLeuProGlnAsnLe 426
 426 LeuSerAspIleuSerLeuGlnSerIleuThrAlaLeuThrIleuGlnValG 443
 556 GCTCTCGACCTTAAGCTCGAGAACCTCAACAGCGTCACTCCAGAGTGG 105
 443 IAsnIleSerAsnThrValGlnGlnLeuLeuLeuSerIysAlaIaVal 459
 106 AAAACATGTCACACCGGACGACAGCTGTCCTGCTCCAAAGTCACTATG 155
 460 ProGlnIysIysGlyAlaIysAsnLeuValSerAspThrProGlnIAsnI 476
 156 CCAAGAAAGAAAGGGGTCAAGAACCTGTGTCACAGAGTCCAGAGCGACA 205
 476 sIysSerIAsnIleCySerProGlnIysSerGlyIysSerAlaGluProA 493
 206 CAATAACCAAGACAGTGAAGCTCGAAGGACAGCGCTACACAGTCAAGCAG 255
 493 IagIyThrProLeuSerGluProProSerSerThrProGlnSerThrHis 509
 256 CCGGACACACCGCTGTGTGAAGCGCCGAGCAGACCGCCACATCCACCAT 305
 510 AlaGluProGlnIAspIAspIysLeuSerGlySerGluAspProLeuG 526
 306 GCGAGACCAACAGGACACTGCTACCTGTAGTGGCTCGAGGACCGCGTAA 355
 526 ValGSerPheLeuThrTyrcysAsnGlnAlaIAspGlySerProIaArgVal 542
 356 GCGCACTGTGCTCAACGACGCCAGCCCGGGGCAACCCCGCCAGGTAC 405
 543 AsnSerAsnSerIysAlaIysProGlnSerValSerThrCySerValTh 559
 406 AACGCACTCCAAAGCTAAGCCCAAGCTGTGTGCACCGTGCCTGTGAC 455
 559 rSerProAspAspMetSerThrIysSerAspAspSerPheGlnSerLeuH 576
 455 CTTCACCTGATGACAGTTCACAGAGTTCGGAAGAGTCTCCAGAGCGTTGA 505
 576 IAspIysLeuProLeuAspSerPheSerIysPheValAlaGlyIAspArg 592
 506 CAGTACCTGGCCCTTGAAGAAGCTCTCCGAATTGTGGCA....AGG 549
 593 AspCySerArgLeuLeuLeuSerAlaLeuAlaGlnIAspLeuAlase 609
 550 AACGGGAAATTCGCGGTGTGTGTCATATGCTCGACAGACAGAAATTCGCCG 559
 609 rGlnIleLeuGlyIleu 614
 600 GGAAT...CTGGGATTTG 612

POSTHAPQEDADILSGSEDPLEBSFLYCNQARGSPAVNSKAKPESVSTCTSPD
DMSTKSDSFOSILSLPDSFKAFAVRGRODPRILLLSALQADLASLDELADLATE
KADKAMAEASIVKSSKPSLENSACIOBSVAKSWPPEPEALPDSLOLDKGN
AKPSFSLPBDVSFAFPPPKKTGTLSPFGKPTLGVAPDPTTAAFCDDPTTAN
SADSANFAMPENI.GDCAKPMGLHPGELTTCBGGKADGSLKIDHEASACIGRO
BEDPGEKVASLPDEKOEYVGLKESQEPGKODLSPKAPLCTKEEVEYVLS
LPLPSTRKEDLEAEFEYSLOELGSEPOGKODLSPKAPLCTKEEVEYVLS
KAGVSCHELSGEVYLLGPTVVESEKVOVSWESLSHMPGEGEGDGEKAVGSTS
DASIAOKPKNPAPVAPPAKKEVPEKSKLSARVNRHLEADSPCRAVLPDOLL
DESCTGPOGOMGAGARGASGLPRKOTSLALSEPDPGPGTLPTTAPDPL
GKORAFKSGKRVKSPKASPSNPALPLVADSPPGSKTEKEDSPSPKODR
SMILRSKTOEIPFSKRRPSSBRLPNKATKILLDSHLPATFVSSSPKGRVS
QARVPKPGAGSLSDRPLHALKRSALFAPVTKRMVLYRSSSSSNAGNGG
KEEPRESPLFKMSSPKAKPKNGEAPKLPPTPACILKLSAALQKAMKT
KYPBRGRLKLEIVQKISPKKFAKAPGAENPLSPISLSDKROGLKAGS
PYVBEGLVAVGKQLPMSGADPLCNPNLSLKMKNKSLSTDCFKTEATSP
FALOGGTALAPKRRKORAGAGLSKGLKEXRYVILPALLTPRBSAGQASD
NSGCKGAKRRKROQVLPDPAPBGRCPQPTKRSKODPHTYNSSTSRKRLTGRKNT
TSPCKGAKRRKROQVLPDPAPBGRCPQPTKRSKODPHTYNSSTSRKRLTGRKNT
PMSGONPAPADARPLCTTDRRAYSAREQOR"

BASE COUNT 1255 a 2038 c 1746 g 876 t

ORIGIN

Query Match 79.1%; Score 4762.2; DB 9; Length 5915;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 4803; Conservative 0; Mismatches 13; Indels 3; Gaps 3;

QY 469 caccagataccagccagtcagtcagtccttcctcgaaagtggttggttcagtcagca 528
DB 449 CGAGATACCAACGACCGAGTCATGCAAGCTTTTCGAGAAAGTGtggTTCATGCGA 508
QY 529 caacagaaactaccagagaccctcgagaaacataccgctcgagaaattacagccagccg 588
DB 509 CAACAAATACCAAGCAAGACTTCGAGAAACATCAAGCTTGAGAAATTAAGAGCAGCCG 568
QY 589 agtcaagccgagctcaagctcgacccgagcagcagcagcagcagcagcagcagcagcag 648
DB 569 AGTCAAGCCGAGCTTAAGCTCGAGCAGCGAGCGCTGCGCAAGACTATTAATTAACCCG 628
QY 649 cagccttaaccagagctatgaggttgagctgagcagccctctgagcagcagcagcagcag 708
DB 629 CAGCCTTACCCGAGCTATGAGGCTGCGCTGACAGCCCTGCGCAAGCTGCAAGCGAGG 688
QY 709 gcgcgcgaacaagtacacagagcagcagcagcagcagcagcagcagcagcagcagcag 768
DB 689 GCCCGGACAAATACCAAG 748
QY 769 agcgcgccttcctcgagctcgagcagcagcagcagcagcagcagcagcagcagcagc 828
DB 749 AGGCGGAGCTTTCCCTGCTAGCGGTGCTCAAGGAGAGAGAGAGAGAGAGAGAGAGAG 808
QY 829 ggtgagagagagcctcaaggtcttgaggggcccaagcagcagcagcagcagcagcagcag 888
DB 809 GGTGAG 868
QY 889 ctaccctcgaggttgcaaatatgatatgagaaactgtagtaaaagacagcagctgcccc 948
DB 869 CTACCTCAAGGAGTGAAGTATGATGAGAACTTGATGAGAAAGAGAGAGAGAGAGAGAG 928
QY 949 agcagcagatgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1008
DB 929 AGCAG 988
QY 1009 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1068
DB 989 CAGCAG 1048
QY 1069 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1128
DB 1049 CAG 1108
QY 1129 cagtcctcccaactcccaactcctcctcgtccagagtggtgagcaggggagc 1188

DB 1109 CAGGCTTCCCAACCTCCCTCCACACTACTCTCTCTCTCCAGAGGTGGGCAAGGAGGAC 1168
QY 1189 caactctataagaggttgacagcagcagcagcagcagcagcagcagcagcagcagcagc 1248
DB 1169 CACTCTATTAAGAGTGAAG 1228
QY 1249 agctccagccttgcccccagggcagcagcagcagcagcagcagcagcagcagcagcag 1308
DB 1229 AGCTCCAGCTGGCCCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1288
QY 1309 ctcaactatgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1368
DB 1289 CTCACACTATGACAG 1348
QY 1369 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1428
DB 1349 CAG 1408
QY 1429 taagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1488
DB 1409 TACGAG 1468
QY 1489 taactacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1548
DB 1469 TACTACAG 1528
QY 1549 ccttcctaaagttcccaacagcagcagcagcagcagcagcagcagcagcagcagcagc 1608
DB 1529 CTTCTCTACAGTCCACAGCAGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1588
QY 1609 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1668
DB 1589 CAG 1648
QY 1669 cccctgtcctatccctcccaagcagcagcagcagcagcagcagcagcagcagcagcagc 1728
DB 1649 CCGCTGTGTAACCTCCCTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1708
QY 1729 tgcagacccctcag 1788
DB 1709 TGCAAGCCCTTCCCAAG 1768
QY 1789 agcctcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1848
DB 1769 AGCTTCAG 1828
QY 1849 ctctccaaagctgtctgtgcagcagcagcagcagcagcagcagcagcagcagcagcagc 1908
DB 1829 CTCTCCAAAGGCTGTGTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888
QY 1909 gaggcagataaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1968
DB 1889 GAGCAG 1948
QY 1969 ggcacacagcgttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2028
DB 1949 GGCACACAGCAGTGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2008
QY 2029 gaggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2088
DB 2009 GAGGCGAG 2068
QY 2089 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2148
DB 2069 CAGGCGCGAG 2128
QY 2149 tccactgttctgtgagactctctcagagcagcagcagcagcagcagcagcagcagcagc 2208
DB 2129 TTCACCTGTGAG 2188
QY 2209 agcctacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2268

4426 atgaagacaaagctgctgacacccggaaggccggggtcgaagctggaagcagctg 4485
 4409 ATGAAGACCAAGGCTGCTGACACCCGGAAGGCGGGGCTGGAAGCTGGAAGCTG 4468
 4486 CAGAAGATCACTCGCCAGCTCAAGAGTTCGATGAAGCGCGAGGCTCTCT 4545
 4469 CAGAAGATCACTCGCCAGCTCAAGAGTTCGATGAAGCGCGAGGCTCTCT 4528
 4546 gtaatacctctgagccacaccccttcgacaaagacccgtggtcgaagctg 4605
 4529 GGAATACCTCTGAGCCACCTCTTCCGACAAAGACCGTGGGCTCAAGGCTG 4588
 4606 agccacagtggtggtggaagaagcctgtaattggtgacacccggcgaagct 4665
 4589 AGCCACAGTGGGGTGAAGAGGCTGGAATGTGGGCAAGGCGCAAGCTCCAACT 4648
 4666 tctgggtctgacgttattcaaaatacaacacacacacacacacacacacac 4725
 4649 TCTGGGGCTGAGTTCGATTAACAAATCCACACAGATCTTAAGGCAATCAG 4708
 4726 aacagtaagaacacgtcttcttactgactgttttaaaacggagcttcacat 4785
 4709 AACAGTAAGAAACGTCTTCTACTGACTGTTTCAAAACGAGGCTTCACATCC 4768
 4786 ggcctcgaagcctggggggaagctggtcgtggtcgaagagagccggaagcga 4845
 4769 GCGCTGCAAGCTGGGGGAGTGCCTGGGGGCTTAAGAGAGGAGCGCGGAGGA 4828
 4829 GGGGCCCATGAGCTCTCAAAAGGCGGCTGAGAGAGGCGGCTTACTGCGGCT 4888
 4846 ggggccaatgactctcccaagagcccgctggaagagcgccatcttggccggc 4905
 4829 GGGGCCCATGAGCTCTCAAAAGGCGGCTGAGAGAGGCGGCTTACTGCGGCT 4888
 4906 ctcttactctcccgagagcagggcagtgacacaaagggcgagtgagacaact 4965
 4889 CTCTTACTCTCCCGAGAGGAGGCGGCTGAGAGAGGCGGCTTACTGCGGCT 4948
 4966 ggaagagagcgaagaagacgaagatgagagagcgtggtcgtcgtcgtcgtc 5025
 4949 GGAAGAGGCGAAGACCAAGAGTGAAGAGGCTGAGGCTGAGGCTGAGGCTG 5008
 5026 aggcctctgcaagccacgaagagagagagagagagagagagagagagagag 5085
 5009 AGGCCTGCGCAACCCCGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 5068
 5086 tattcaagacggaagcgtcctcactcgtggcggtggtggtggtggtggtggt 5145
 5069 TATTCAGAGCGGAAGCCCTCACTCGGGCGGCGCAAGAGAGAGAGAGAGAG 5128
 5146 aaggggctgcaagcgaagcgaagcgaagcgaagcgaagcgaagcgaagcga 5205
 5129 AAGGGGCTGCGCAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5188
 5206 gaaatccgctcaagtaattctctctctgcaagcgtgagtcgaagcgaagcga 5265
 5189 GAAATCCGCTCAAGTACATTCTCTTGAAGGCGGTGAGTGAAGAGAGAGAGAG 5248
 5266 ggccttcaacccctcgtgc 5284
 5249 GCCTTCTCAACCCCTGAGAGC 5267

RESULT 5
 LOCUS AX024201 5395 bp DNA
 DEFINITION Sequence 1 from Patent WO0008143.
 ACCESSION AX024201
 VERSION AX024201.1 GI:10184512
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 5395)

AUTHORS Poustka A., Lehrach H., Radelof U. and Seranaki P.
 TITLE Gene isolated on the short arm of human chromosome 17
 JOURNAL Patent: WO 0008143-A 1 17-FEB-2000;
 DEUTSCHES KREBSFORSCH (DE); POUSTKA ANNEMARIE (DE); LEHRACH HANS
 (DE); RADELUF UWE (DE); SERANSKI PETER (DE); MAX PLANCK INST
 FÜR MOLEKULAR (DE)
 FEATURES
 source 1..5395
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..>5394
 /note="unnamed protein product"
 /protein_start=1
 /protein_id="CAC08943.1"
 /db_xref="GI:10184513"
 /translation="MOSRRERCGFHGKNOONVOOTSEHSEUENVRPOPSAGLSGDR
 ILADYNNPQPSYSGAGAGPSCTGAATAAATKTRSKALPTQGLQRRAPFGTY
 QDSPTFRAGREESIQMGAPPPPPQPLPAVAATIDENLMTKTAAPVPSRYAQ
 GAQVFTSHLVQPPPPQPLAVPLQRLQRLDNLASPLPPQGMAGVSHYSCOT
 APTAPDRPLTSSSLAPGQVONLHAYOSGRSLTYDQDQDQDQDQDQDQDQDQDQ
 AQTALYONLAKYOHGOOGGYCOEDPAVVRPEGYOTFSPSSSPARSVGRPEY
 SETSPMLPMLNFPYSOQPLSTGAPAGITPHSHFPLNPPPTATSPADLALT
 SOVENTISNTVOULLSKAIVOKKGVKNVSTPPQHSOHCPSBSCSAPACTEL
 SEPSSTPOSTHAPROEDYLSSEDEPLEKSLTYCNQKNSPARVNSKAPESVST
 CSVTSPPDMSKSDSFQSLGSLPDSFSKTVAGEROCPRLLSLAQEDLASLIG
 LOELIGKADKAWAPSLVSDSKPEPFLKSHSACLSYAKSAMPRPCEPLASLIL
 QLDGASNAKDFSPGLFEDCVAFATPDCKRGLTSPFCTKTLGVAPAPPTAAVDF
 PPTAASASDANPFWMEENLADGCRNGLRPGHLLTGLEBOGKASDGLSTGOTHR
 SAGICRPEDEPCEKVAALPDPEQEVGVKVEAGGLQCEPYAADNWLDESHHC
 STAPGDLPLPPTSKEDLEHERSTSCLELSSQEDRGMDPLSPALPTKEE
 VEELIDSKRAGWSPOHLSGESVILLAPTYTGLSKVSPFSDSSLSHKKDEEPGE
 STAEVDSKALQKRNKPAVEAIKAKDPVVRGKLSRVRGLPEAESQCAAY
 LPKDLPLPESCTPGQGEAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
 PAPDKLGKORAPSKRVKSPSPAASSNPALPYASDSSPMGSKTENDSSG
 TPBGDQSMILRSRMTQTEHRSRRPSEBGLPBCRTKTLDNHSLATRYKSSD
 QKEATVQRARVTRKPAAGSKISDPRHLAKRSAMAVPFRKRVLSNRSSSSNA
 SONGDEKERPGSPFLFKEMSSPKNAPRTGNSEPTKLLPPEPDCILKLSRA
 FQGMKTRVLPVRKRGSLKLAIVQITSPSLKPKACAPASGMDPLSPSTKDNG
 LKAGGSPVVEBGLVNYGTGOLPTGADPLCRNPTYRSIKGLNNRASGTGASD
 NSGGGKPKMEELGLASOPPEGRPOPTRAQKQPGHNTSSYSKRLPLGAKNT
 TSPCKGRARRROOVLPDAPRTKRLYSCKRLRDSSTPRAPSPVAYEKRA
 FTTCTVYNSRGOARPRHRSKSSASSSSSSFLDAGASLTPDGSLIIPPSL
 PLSTMLGAVYSKALSTSCVLCQNPANRKGDLGDCGYTPPEHCLPKPKPKRL
 VPEBCTEASLPLELTLKPECAAAATAPKPRPDVPAPOGRLTASAKSKRL
 QSCYCDGREDGELEAPADGHNVAHCAVYTGCVTVAGKLGLEQEMKAVADMG
 GTVALAGDESLPGLRFAFLPQSPWDCVLAATSPSMBRCVPAKKRPGLAAT
 KDAAPTTTRVPM"

BASE COUNT 1117 a 1877 c 1595 g 806 t
 ORIGIN
 Query Match 76.8%; Score 4627.6; DB 6; Length 5395;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 5131; Conservative 0; Mismatches 9; Indels 398; Gaps 10;
 QY 490 atgacgctcttcgagaaagctggtttccatgycacaaacgaactacacagcagc 549
 Db 1 ATGACGCTCTTTCGAGAAAGGTGTGTTCCATGCAACAAAGAACTACACAGAAC 60
 QY 550 tcgcaggaacacacgcgcctagagattcagagcagccgagtcgagcggtacgtgc 609
 Db 61 TCGAGGAAACATCAGGCTTGAATTCAGAGCGAGCGAGTCAAGCTCGGCTAAGCTGC 120
 QY 610 gaccgagcagcgtcgtcgcgaagagcctttaacccgagcgttaacccgagctatgag 669
 Db 121 GACCGGAGAGGCGTGTGCTGCAAGAGCTATTATACCCGACAGCTTACCGAGCTATGAG 180
 QY 670 ggtggcgctggcagccctctggaactgagcgcggtggtgcgcgcagacaagtacaccca 729
 Db 181 GGTGGCGCTGGCAGCCCTCTGCACTGCAAGCGGCGGCGCGCAGCAAGATACACCGA 240
 QY 730 ggcagcaagcgccttcacacagcagcctcgaggggagcggtcttcctgcgtac 789

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:

December 4, 2001, 16:09:27 ; Search time 4574.73 Seconds
(without alignments)
21716.258 Million cell updates/sec

Title:

US-09-508-821B-5

Perfect score:

6022
1 ggcacgacgagcccaaggy.....ggagctcttttggtgcag 6022

Sequence:

IDENTITY NUC
Gapop 10.0, Gapext 1.0

Scoring table:

1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters:

2944280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

GenBank: *
1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vl:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.om:*
20: em.ov:*
21: em.pat:*
22: em.ph:*
23: em.pl:*
24: em.ro:*
25: em.sts:*
26: em.sy:*
27: em.un:*
28: em.vl:*
29: em.yl:*
30: em.htgo.hum:*
31: em.htgo.in:*
32: em.htgo.in:*
33: em.htgo.hum:*
34: em.htg.in:*
35: em.htg.ro:*
36: em.htg.other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	6022	100.0	6022	6 AX004680	AX004680 Sequence
2	5961	99.0	160544	2 HS253P07	AL354000 Homo sapi
3	5901.2	98.0	23843	9 HBA271791	AL271791 Homo sapi
4	4762.2	79.1	5915	9 AB058723	AB058723 Homo sapi
5	4627.6	76.8	5395	6 AX024201	AX024201 Sequence
6	4627.6	76.8	5667	9 HBA271790	AJ271790 Homo sapi
7	3664	60.8	7222	10 MUSMA	D29801 Mouse mRNA
8	970.4	16.1	2660	9 HSM801433	AL133649 Homo sapi
9	495	8.2	585	11 AF021115	AF021115 Homo sapi
10	391.8	6.5	495	11 AF021118	AF021118 Homo sapi
11	326	5.4	542	11 G09710	G09710 human SFS C
12	317	5.3	330	6 AX004677	AX004677 Sequence
13	206.4	3.4	328	6 AX004676	AX004676 Homo sapi
14	110	1.8	6542	9 AB006630	AL031346 Human DNA
15	110	1.8	43761	9 HS9782	AY007594 Mus muscu
16	101.2	1.7	6212	10 AX007594	U70655 Mus muscu
17	85.4	1.4	1512	10 MM070656	U70656 Mus muscu
18	85.4	1.4	1512	10 MM070656	X67204 Mus muscu
19	85.4	1.4	14625	10 MMSRLC	AC084799 Mus muscu
20	83.6	1.4	303091	2 AC084799	E11536 DNA from y
21	82.2	1.4	10266	22 E11536	AC005918 Homo sapi
22	82	1.4	87076	9 AC005918	AC087563 Homo sapi
23	81	1.3	298166	2 AC087563	AL137857 Homo sapi
24	79	1.3	152547	2 AL137857	AC073946 Mus muscu
25	78.8	1.3	197232	2 AC073946	U70651 Mus muscu
26	78.6	1.3	1611	10 MMU70651	AL157736 Homo sapi
27	78.6	1.3	200542	2 CNS01R63	AC007263 Homo sapi
28	78.4	1.3	167390	9 AC007263	U70657 Mus muscu
29	76.4	1.3	1549	10 MMU70657	AC021049 Homo sapi
30	75.8	1.3	178343	2 AC024049	U66463 Mus muscu
31	74.8	1.2	1968	10 MMU46463	AC000097 Homo sapi
32	74.4	1.2	162269	9 AC000097	AC007663 Homo sapi
33	74.4	1.2	168239	9 AC007663	AC006549 Homo sapi
34	74.4	1.2	174840	9 AC006549	AC006547 Homo sapi
35	74.4	1.2	179269	9 AC006547	AL049849 Homo sapi
36	74.2	1.2	41185	9 HSB11C6	AL161725 Human DNA
37	74.2	1.2	181179	9 AL161725	AC044845 Mus muscu
38	73.8	1.2	235852	2 AC044845	U70649 Mus muscu
39	73.6	1.2	1547	10 MMU70649	U70642 Mus muscu
40	73.4	1.2	1546	10 MMU70642	U70645 Mus muscu
41	73.4	1.2	1546	10 MMU70645	U70641 Mus muscu
42	73.4	1.2	1550	10 MMU70641	U70650 Mus muscu
43	73.4	1.2	1552	10 MMU70650	AF337043 Mus muscu
44	73.4	1.2	2298	10 AF337043	AF337045 Mus muscu
45	73.4	1.2	2319	10 AF337045	

ALIGNMENTS

RESULT	1	LOCUS	AX004680	6022 bp	DNA	PAT	24-AUG-2000
DEFINITION	AX004680	Sequence 5 from Patent WO915639.					
ACCESSION	AX004680						
VERSION	AX004680.1	GI:9928116					
KEYWORDS	unclassified.						
SOURCE	unclassified.						
ORGANISM	unclassified.						
REFERENCE	1 (bases 1 to 6022)						
AUTHORS	Rouleau G.A. and Jooper R.						
TITLE	Polymorphic cag repeat-containing gene and uses thereof						
JOURNAL	Patent: WO 915639-A 5 01-APR-1999;						
FEATURES	LOCATION/Qualifiers						
SOURCE	1..6022						
	/organism="unclassified"						
	/db_xref="taxon:32644"						

BASE COUNT 1247 a 2060 c 1799 g 916 t

ORIGIN

Query Match 100.0%; Score 6022; DB 6; Length 6022;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6022; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcagagcccaaggaggtgagggagcgaaattctctgtctaaatgcttttgagc 60
DB 1 GGATCCAGCAGCGCCCAAGGGGATGAGGAGCGGAAATTCCTGCTAAATGCTTTGAGC 60
QY 61 tgcacgaagggctggagtgatgggtgggagacattgggagagctggcaatggcg 120
DB 61 TGCAGGAAGGCTGGGAGTGAATGGGTGGGAGATGGGAGAGCTGGCAATGGGCG 120
QY 121 gggggggcggtgagctcccaagtagctggcgctggcgagcggttttggctcccgat 180
DB 121 GGGGGGGGCGGTAGCTCCCAAGTACCTGGCGTGGGAGCGCGGTTCCTCCCGCAT 180
QY 181 cagtggccgtctctggcaagctcagctcgaggcgatgtgggagcggaattacaagca 240
DB 181 CAGTGGCCGCTCTTGGCAAGACTCAAGCTGCAGGCGATGTGGAGGGGGAATTACAGAGCA 240
QY 241 caactccctgacacagaagttgcaatatgcacagctggtgggagctcagcgcaag 300
DB 241 CACTCCCTGACACAGAAGTTGCAATATGCGCAAGCTGGTGGGAGGCTTCAGGGGAAG 300
QY 301 gggggactaataagctgcggcggggagcagcgaggtgggaggtgggaggggtg 360
DB 301 GGGGGACTAATAAGAGCTGCGCGGGGAGCAGGAGTGGGAGTGGGTGGGAGGGTG 360
QY 361 cttcttgaggaagaagtgcccgctgataatgcctatctctctgcccctccctgcc 420
DB 361 CTTTCTGAGGGCAAGAAAGTGGCGCGTGAATGCTCATCTCTGCCCCCTCCCTGCC 420
QY 421 catctccctccctccctccctccctccctccctccctccctccctccctccct 480
DB 421 CATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 480
QY 481 gcccgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcag 540
DB 481 GCCCGAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 540
QY 541 cagcagacctcgaggaagaacatcacctctagagattacaaggcagcgagtcagcg 600
DB 541 CAGCAGACCTCGAGGAACATCACCTCTAGAGATTACAGGCGCGAGTCAGGCGCGG 600
QY 601 ctaagctgagcgagcgagcgctgctcgcaaggactattataacccgagccttacccg 660
DB 601 CTAAGCTGAGCGAGCGGAGCGGCTGCTCGCCAGGAGCTATTATACCGCGAGCTTACCG 660
QY 661 agctatgaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 720
DB 661 AGCTATGAGGTTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGG 720
QY 721 tacacacgaggaaggaagccctgcccacacagaaagccctgaggggagcggttttc 780
DB 721 TACCACCGAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
QY 781 cctggttaagcggttcaggaacagcagcagcagcagcagcagcagcagcagcagc 840
DB 781 CTTGCTTACGGGCTCCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840
QY 841 cttcaggtctgg 900
DB 841 CTTTCAGGCTTGG 900
QY 901 gtggcgaagtatgatgagaactgatgaaaaagacagcagtgccccagcagcgatgat 960
DB 901 GTGGCCAAAGTATGATGAGAAGTGTGATGAAAAGACAGCAGTGGCCCCCAGCAGCAT 960
QY 961 gpadgagcgagggcgaggtgccccttcgactcactccctgacgtccagcagccacog 1020

DB 961 GCAGAGCAGGGGCGCCAGGTGCTTTCGGACTACTCCTCTGCACGCTCAGCAGCGACCG 1020
QY 1021 cggccccagcagccctggcatacccaagctcccaaggcagaagctgcagaacgacatt 1080
DB 1021 CGGCCCCAGCAGCCCTGGCATACCCCAAGCTCCAAAGGCAGAAGCTGCAGAACGACATT 1080
QY 1081 gctccctctgctccctcccaagggtacacacttctcagcattccagtccttcccc 1140
DB 1081 GCTCCCTCTGCTCCCTTCGCCAGGTACCCACTTTCCTCAGCATTCGACATTCGAGTCCCTTCCC 1140
QY 1141 ecctcctccactactcctctctcaggggtggtgggaggggggggggggggggggg 1200
DB 1141 ACTCTCTCCACTACTCTCTCTCTGTCAGGGTGTGGGAGGGGGGGGGGGGGGGGG 1200
QY 1201 agttgacagcagccagctgcccagggccccatgacagggcgctgactgccaagctccagctg 1260
DB 1201 AGTTGCACAGCAGCCAGCTGCCAGGCCCATGACAGGCGCGTACTGCCAGCTCCAGCGCTG 1260
QY 1261 gccccggggcagcggttcagaatcttcctacagtcggggcgccgctcagctatgac 1320
DB 1261 GCCCGGGGCGAGCGGTCCAGATTTTCATGCTTACCATGTCGGGGCGGCTCAGCTATGAC 1320
QY 1321 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380
DB 1321 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1380
QY 1381 catgcccaggaacccctccattaccaaaaacctgccaagttatcagcaactaaggggcagcaa 1440
DB 1381 CATGCCAGGAACCCCTCCATTACAAAACCTCGCCAAAGTATCAGCACTACGGGCGAGCAA 1440
QY 1441 ggcagggctactgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcg 1500
DB 1441 GGCAGGGCTACTGCGCAGCGGAGCGAGCGGCTCCGAGCCCCCAGAGAGCTACTACCGAGCC 1500
QY 1501 ttccagccccagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560
DB 1501 TTCAGCCCCAGCTCCAGCCACTCACCGCGCGCTCGGTGGCGCGCTCACCTTCTCTACAGT 1560
QY 1561 tccacacgctgcgctgatatccaaacctggagaactttccctacagcgagcgagcgctc 1620
DB 1561 TCCACACGCTGCGCGCTGATGCCAAACCTGGAGAACTTTCCCTACAGCGAGCGCGCTC 1620
QY 1621 agcagcgggccttcccgagcgagcagcagcagcagcagcagcagcagcagcagcagcag 1680
DB 1621 AGCACCGGGGCTTCCCGCAGGGATCATGACACAGCCACTTCAAGCCCTGCTCAAT 1680
QY 1681 cctccccaagcagatgcacacagctctgtggacacccagggctggcaactgcaagccccctt 1740
DB 1681 CCTCCCAACGAGTGCACACAGCTCTGTGGACACCCAGGCTGGCAACTGCAAGCCCCCTT 1740
QY 1741 cagaaggaacagctccctgagaacctgctgtcgatctcagcctgagagcgctcagcgcg 1800
DB 1741 CAGAAGGACAGCTCCTGAGAACCTGCTGTGGATCTCAGCCTGCGAGAGGCTCACGGCG 1800
QY 1801 ctgaccttcagctgagaacatctccacacgctccagcagctgctgctctccagcct 1860
DB 1801 CTGACCTTCAGGTGAGAACATCTCCACACCGCTCCAGCAGCTGCTGCTCTCCAGGCT 1860
QY 1861 gctgtgcgcaggaagaaagtgctcaagaacctgctgtccaggacccagcagcagcataaa 1920
DB 1861 GCTGTGCGCGCAGGAAGAGTGTCAAGAACCTGCTGCCAGGACCCCGAGAGCAGCATAAA 1920
QY 1921 agcaggaactgcagcccccgagggagcggtactcagcgagcgagcgagcgagcgagcgctg 1980
DB 1921 AGCCAGCACTGCGAGCCCCGAGGGAGCGGCTACTCAGCCGAGCGCCGAGGACACCCGCTG 1980
QY 1981 tcagagcgccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2040
DB 1981 TCAGAGCGCGCGAGCAGCAGCGCCACAGCTCCACAGTCCGAGCGCCGAGAGGCCGAGCTAC 2040
QY 2041 ctgagcggtcccgaggaacccactgagcgagcttctctactgcaaccagggccggtggc 2100
DB 2041 CTGAGCGGCTCCGAGGACCCACTGGAGCGCAGCTTCTCTACTGCAACAGGCGCGTGGC 2100

QY	2101	agccctccaggggttaaaagaaactctgaagggcagaagccgaatccggtctcaactcttct	2160
Db	2101	AGCCCTCCAGGGGTAAAGAAACTCTGAAGGGCAGAGCCCAAGTCCGATCCACCTTCCT	2160
QY	2161	gtgaaacctctctgaagacatctccaccaatctagaacactctctccagaagctcaacgycg	2220
Db	2161	GTGAACCTCTCTGAAGACATGCTCCACAACTGACACTCTTCCAAAGCTTCAGAGCTTCACGGC	2220
QY	2281	ctgactcaagggccctctgacagagagacccctggccctgcagagatcccttgaggcttgagagacgc	2340
Db	2281	CTGACTCAGGCCCTCTGACAGAGAGACCTGGCCCTTCAGACTCTTGAGGCTTCAGAGAGCC	2340
QY	2341	atcsgtgaagaagcgcgaacaagcttgggtctgaagacaccagccctggtcaagaagcaagc	2400
Db	2341	ATCGGTGAGAAAGCCGACAAAGCTTGGGTGTAAGTACCCAGCTGCTGACAAAGACACAGC	2400
QY	2401	aagcgacaacctctgcctgctgagaaacacagccgcctgccttgagacctctgtgccaagaatgag	2460
Db	2401	AAGCGACACCTCTGCCTGCTGAGAAACACAGCCGCCTGCTGAGACCTCTGTGCCAAAGATGAG	2460
QY	2461	tggccccggcctctgggagacccgagagccctgcgcagctcctctgacgttgagcaagagggcggc	2520
Db	2461	TGGCCCCGGCCTCGGAGAGCCGAGAGCCGCMGCGCCGACTCTGTGACGTGAGACAAAGCGGCG	2520
QY	2521	aatgccaaggaactcaagcccagggctggtttgaagacccttcggtgcttcgtctagacct	2580
Db	2521	AATGCCAAGGACTCAAGCCAGGGCTGTTTGAAGACCTTCCCTGAGCCTTGCTAGAGCT	2580
QY	2581	gaccccaaaaagacaactcgtgctccctccctctctgtgtacaaagcccaacctctgggttctct	2640
Db	2581	GACCCCAAAAAGACAACTGCTCTCTCTCTTGGTATCAAGGCCACCCCTTGGGGTTCTCT	2640
QY	2641	gctccagaaccccaactacagagagcttlttgacgtgtttcccgagaaacaaaccgctgcacagatca	2700
Db	2641	GCTCCAGACCCCACTACAGAGAGCTTTLTGACGTGTTTCCCGAGAAACAAACCGCTGCACAGTCA	2700
QY	2701	ggggagacagcgcaaaccccttgcctctggccagagagaaacacttgaggatgagcttgcacagag	2760
Db	2701	GGGAGACAGCGCAAAACCCTTGCCTCTGGCCAGAGAAACACTTGGGATGAGCTTGTGCCAAG	2760
QY	2761	tggggaattgcacacctctgcgagacttaacaaagggcccctgtagcagagtggtggaaagccctacaat	2820
Db	2761	TGGGGAATTGCACACCTCTGCGAGACTTAACTAAAGGGCCCTGAGAGCGGTGGGAAGCCTCAGAT	2820
QY	2821	ggatcatcagaaggggagacaaccatgtaggcttcgagctgcctgagcttcacagagagagagac	2880
Db	2821	GGATCATCAAGGGGAGACAAACCATGTAGGCTTCGAGCTGCCTGAGCTTCACAGAGAGAGAGAC	2880
QY	2881	cccacctgggagaaagtgagctcgttgccccggggagacttaacaaagagagagagtggtgtgtg	2940
Db	2881	CCCCCTGGGAGAAAGTGAAGTGGCTCTTTCGCCGGGAGACTTAACTAAAGAGAGAGAGTGGGTGG	2940
QY	2941	gt'gaaagagagagcaggtgaggtcgtgcgtgcagtgagcccgagagtggtgcagaggtgcacggtgtg	3000
Db	2941	GT'GAAAGAGAGAGCAGGTGAGGTGCTGCTCAAGTACCCCGAGGTGACCAAGGTGCAGCCGGTGG	3000
QY	3001	ctggagagacagccggagacatcgtctgtccacacggcagactctggggagactccaactcctctgagca	3060
Db	3001	CTGGAGAGACCGGAGACATCTGCTGTCTGTCCACACGGCAGACTCTGGGAGACTCCCACTCTGAGCA	3060
QY	3061	cccacacagcagagagagagacacttgaaagcttgagggagaggtatccctacatctctgtgagcttc	3120
Db	3061	CCCAACACAGAGAGAGAGACACTTGAAGCTTGAAGGAGAGGTATCCCTACATCTCTGTGAGCTTC	3120
QY	3121	ctggggagcccccgagagagagcctcgtgatagtgcagagcccgagctgtcacaaccagaagccccacatc	3180
Db	3121	CTGGGAGACCCCGAGAGAGAGCCTCGTGAATGTCAGAGACCCGCTGTACCCCAAGCCCCACATCT	3180

[illegible]

Db 4261 GGAAGGAGGAGGAGGCTGAGGGTTCCTCCACACCTCTTCAAGAGGATGTCTTCTCCCAAGA 4320
 QY 4321 aagccaagcccaagggcaatggagcctgcacaaagctcccaaccccgagaccc 4380
 Db 4321 AAGCCAAAGCCCAAGGCAAGGGAGGAGGCTTGCACAAAGCTCCCAACCCCGGAGACCC 4380
 QY 4381 caattcctgctcaagctgcgctctggcagccttccagggggggccatgaagacaaagt 4440
 Db 4381 CCATTCCTGCTCAAGCTCGCTCTCGGAGCCTTCCAGGGGGCATGAAGCAAGGTG 4440
 QY 4441 ctgcaccccggaagggcgccctgaagctggaagccatctgcaagatcaactcg 4500
 Db 4441 CTGCCACCCCGAAGGGCGGGGCTGAAGCTGGAAGCCATGCTGCAAGATCACCTCG 4500
 QY 4501 ccagcctcaagaagtgcgcatgtaagggcgagggcctctcctgttaactctgagc 4560
 Db 4501 CCCAGCTCAAGAGTTCGATGTAAAGCCGACAGGGGCTCTCCTGGTAATCTCTGAGC 4560
 QY 4561 caatcccttccgacaaagacgtgggtctaaaggggtgctggggcgagggggggg 4620
 Db 4561 CCATCCCTTCCGACAAAGACCTGTGGCTCAAGGGTGTGGGGGCGAGCCAGTGGGGGTG 4620
 QY 4621 gaagaagcctgtaaatgtggcagcggggcgaaagctcccaacttctgggctgatccg 4680
 Db 4621 GAAGAAGGCTGTAAATGTGGGCAACCGGCGAGAGCTCCCACTTCTGGGCTGATCCG 4680
 QY 4681 ttatcagaataccaacaaagatccttaagggcaaaactcatgaacagtaagaactg 4740
 Db 4681 TTATGCAATCCACCAACAGATCTTAAAGGCAAACTCATGAACTAGAAACTG 4740
 QY 4741 tcttactgactgtttcaaaacagggccttcaatcccgaggcctgagcctggg 4800
 Db 4741 TCTTCTACTGACTGTTCAAAACGAGGCTTCATATCCCGAGGAGCCCTGCAGCCTGGG 4800
 QY 4801 ggactgctgcctgctaaagaagcagcagcagcagcagcagcagcagcagcagcagc 4860
 Db 4801 GGGACTGCTGCTGGCGCTTAAGAAGAGAGAGCGGAGAGCGGGGCGGCGGCGGCGGATG 4860
 QY 4861 tccaaagggcgctggagagggcgcttctggcgggctctgctcctgactcccgga 4920
 Db 4861 TCCAAAGGCGGCTGGAGAGGCGGCTATCTTGGCGGCTCTGCTGACTCCCGGA 4920
 QY 4921 gacagggcagctggcacaaagggcgagtcagtgaggacaaactctgtgagggaggaag 4980
 Db 4921 GACAGGGCAGTGGCACACAGAGGGGCGAGTGAAGCAAACTCTGCTGAGAGGAGCAAG 4980
 QY 4981 ccaaatgagagagctggcgctgcctccagcccgagggcgagcctgcagccg 5040
 Db 4981 CCAAAAGATGGAGAGTGGGCGCTGCTCCAGCCCGGAGGCGGCGGCGGCGGCGGCGG 5040
 QY 5041 cagcaagggcacagaaacagcagcagcagcagcagcagcagcagcagcagcagcag 5100
 Db 5041 CAGCAAGGGCAGAAAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5100
 QY 5101 cgctcactcgggcgggcggaagaaacacacacacacacacacacacacacacacac 5160
 Db 5101 CGCTCACTCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5160
 QY 5161 cgacagacaaacagcagctgctgctcctgagctccgagagcctgaatccgctcaag 5220
 Db 5161 CGACGACGACACAGCAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5220
 QY 5221 tacattctcttgaagcgctgaggtgagcagcagcagcagcagcagcagcagcagc 5280
 Db 5221 TACATTTCTTTGCAAGCGGCTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 5280
 QY 5281 gtgagggtggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5340
 Db 5281 GTGCGGGTGGAGAGCGGAGAGCGGCTTCAACCACTATGCACTGTGTCACTCCCTCG 5340
 QY 5341 gatgcgccaagcccaaggaagccttctctctctctctctctctctctctctctct 5400

Db 5341 GATGCGCCCAAGCCCAAGAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5400
 QY 5401 tcttctctctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5460
 Db 5401 TGCTTCTCTTGGATGACAGCGGGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5460
 QY 5461 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5520
 Db 5461 CAGCCGCGGCGCT 5520
 QY 5521 gacctgagctact 5580
 Db 5521 GCGCTGAGTACT 5580
 QY 5581 ctgggggacctctgtggggccttactacacctgaacctgctctctctctctctctct 5640
 Db 5581 CTTGGGGAGCTCTCTGGGCGCTTACTTACTTGAACACTTCTCTCTCTCTCTCTCTCT 5640
 QY 5641 ctcaaggaagagtgaggccaaagaagcagcagcagcagcagcagcagcagcagcagc 5700
 Db 5641 CTCAGGAGAGAGGTGGGGCGGAGAGGCACTGTGAGAGGCTCTCTCTCTCTCTCTCT 5700
 QY 5701 acaactcaaggtcccgagtgctgagcagcagcagcagcagcagcagcagcagcagc 5760
 Db 5701 ACATCAAGGTCTCCGAGTGTGAGCTTCCGCGCACTTCCGCGGAGGCGGCGGAG 5760
 QY 5761 ccagctgaccccgccaaagcagcagcagcagcagcagcagcagcagcagcagcagc 5820
 Db 5761 CCAGCTGACCGCGGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5820
 QY 5821 gctcagagctgctactgctgctgctgctgctgctgctgctgctgctgctgctgct 5880
 Db 5821 GCTGCAAGAGCTGCTACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5880
 QY 5881 cgacaaggtgcgaacatgagtcagcagcagcagcagcagcagcagcagcagcagcagc 5940
 Db 5881 CGACAGGCTGCGAACAATGATGAGTGCAGCAAGAGGCTTCCGCGGAGGCGGCGG 5940
 QY 5941 ccaggagcactgggtgctgctgctgctgctgctgctgctgctgctgctgctgctgct 6000
 Db 5941 CCAGGAGCACTGGTGTGATGAGGCTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 6000
 QY 6001 cgggaagctcttgggtgctgag 6022
 Db 6001 CGGGAAGCTCTTTGGGCTGCAG 6022

RESULT 2
 HS253P07/c
 LOCUS 17 clone PAC RPI-1 253P07 map 17p11.2, ***
 DEFINITION SEQUENCING IN PROGRESS ***, 2 ordered pieces.
 ACCESSION AL354000 AL035362
 VERSION AL354000.6 GI:13752103
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 160544)
 AUTHORS Ramser,J., Langer,I., Steffens,C., Klein,M., Borzym,K., Junker,E.,
 Marquardt,I., Moll,K., Radelof,U., Francis,F., Serauski,P.,
 Poustka,A., Reinhardt,R. and Leirach,H.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 160544)
 AUTHORS MOLGENR.
 TITLE Direct Submission
 JOURNAL Submitted (28-APR-2000) MOLGENR, Abt. Leirach, Max Planck Institut
 fuer Molekulare Genetik, Innestrasse 73, Berlin, 14195 Germany
 COMMENT On May 14, 2001 this sequence version replaced gi:9931105
 gi:10280527.
 contig 01 1. 86093
 contig 02 86194. 160544

Clone received from the Resource Centre of the Human Genome Project
at the Max-Planck-Institute for Molecular Genetics.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

1 86093: contig of 86093 bp in length
* 86094 86193: gap of 100 bp
* 86194 160544: contig of 74351 bp in length.

Location/Qualifiers

1. 160544
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17p11.2"
/clone="PAC RPCI-1 281113"
/clone="PAC RPCI-1 253P07"
/clone.lib="RPCI1,3-5 Human PAC library, originating
institute: Roswell Park Cancer Institute, creator: Pieter
de Jong, P. Ioannou"

misc_feature

1
/note="region between markers D17S842-D17S953"

misc_feature

160544
/note="SP6_end:PAC RPCI-1 281113"

BASE COUNT 35434 a 44179 c 44032 g 36798 t 101 others

ORIGIN

Query Match 99.0%; Score 5961; DB 2; Length 160544;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 6012; Conservative 0; Mismatches 10; Indels 5; Gaps 4;

QY 1 ggtatccagagagcccaagggatgagggagcgaattcctctgtaattgtttgagc 60
Db 110339 GGATCCAGAGAGCCCGAGGATGAGGAGGCGGAAATGCTGTGAATGCTTTGACC 110280
QY 61 tctcagaagagagctgagatgagtgaggagacattgagagagagctggaatggcg 120
Db 110279 TCTCAAGAGAGGCTGAGAGTGAAGGTGGGAGACATTGGGAGAGAGCTGCAATGGGCGG 110220
QY 121 ggggggggggggtagctcccaagtgactggcgctggagacgggttttgcctccgcat 180
Db 110219 GGGGGGGGGGGTAGCTCCCAAGTGAAGTGGCGCTGGAGCGCGTTTGGCTCCGCAAT 110160
QY 181 cagtgagcgtccttggaagactcagctgcagggagatgagggaggaattaaagagca 240
Db 110159 CAGTGGCGCTTGGCAAGCTGACAGCTGAGAGTGGAGGAGGAATTAAGAGCA 110100
QY 241 caccctcgcagacagaagtgatgcaatgagcagcagctgtgaggaaggtcagggcgaag 300
Db 110099 CACTCCCTGACAGAGAGTGTCAATATGCGCACACACTGATGAGGAGAGCTCAAGCGAAG 110040
QY 301 gggggagactaataagagctgcgggggagcaggaaggtgaggaaggtggtggaggggtg 360
Db 110039 GGGGAGACTAATAAGAGCTGCGGGGAGCAGGAGGTGGGAGGTGAGTGGAGAGGTG 109980
QY 361 cttctcgaagcaaaagagagtgccgtgttaattcgttaactctgcccctccagcc 420
Db 109979 CTTTCTGAGGCAAAAGAGAGTGGCCGTGTGAATGCTCACTCTGCCCCCTCCCTGCGC 109920
QY 421 catctccctccctcctcctccctccctccctccctcttttcttttcttaagaataacca 480
Db 109919 CATCTCCCTCCCTCCTCCTCCTCCTCCTCCTCCTCCTTTTCTTTTCAAGATAACCA 109860
QY 481 gcccgagtaagagcttcttcgagaaggtggtttcattcattgcaacaacaactaac 540
Db 109859 GCCCGAGTATGACAGTCTTTTTCAGAGAAAGTGTGTTTCATGGCAAAACAACTAC 109800

QY 541 cagcagacctgcaggaatacatcagcctatagaatataagcagcgagtcagcgagg 600
Db 109799 CAGCAACCTCGCAGAGAAATATACGCTATAGAAATTAAGGACACGAGCGAGCGGG 109740
QY 601 ctaagctgagacgagcagcggtctctgcaggaaggaattatataccgagcctaccg 660
Db 109739 CTAAGCTGAGACCGGACAGCGGCTGCTCCGCAAGAGCAATATATACCCGACCTTACCG 109680
QY 661 agctatagaggtggtgctgcaagccctctgctcagcctcagcggtgagcgagcaag 720
Db 109679 AGCTAAGAGGAGGAGGCGGTGGGACGCTCTGACCTGACCTGAGCGGTGGCGGAGAG 109620
QY 721 tacacagagagcagaagagccctgcccacacagaagagcctgagggagggcgcttcc 780
Db 109619 TACCAACGAGGAGCAAGGCGCTGCCCCACACAGAGCGCTGAGGGAGGCGGCTTTC 109560
QY 781 cctgtcagaggtgctcaggaagcagccctaccccaagcgctatgctgtgagagagc 840
Db 109559 CTTGCTACAGGCGTCCAGAGACAGACGCCCTTACCCAGGCGCTATGCTGTGAGGAGAGC 109500
QY 841 cttaaggttgg 900
Db 109499 CTTGAGGCTTGG 109440
QY 901 gtggcagaatgatatgataaactgatatgataaagacagcagtgcccccagagagagat 960
Db 109439 GTGGCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 109380
QY 961 gcaagagaggg 1020
Db 109379 GCAAGAGAGGG 109320
QY 1021 ccggcccaagagagccctctgctcctcccaagcctcacaagagagagagcatt 1080
Db 109319 CCGCCCAAGAGAGCCCTCTGCAATACCCAGCTCCAAAGAGAGAGAGTGTGAAGAGAGAT 109260
QY 1081 gctccctctgacctcccccagaggttaccacattctccagatccagctccctcc 1140
Db 109259 GCTCCCTCTGCGCTTCCCGCCAGAGGATCCGCTTCTGAGATTCAGAGCTTCCCG 109200
QY 1141 acctctccactact 1200
Db 109199 AACTCTCTCACTCACTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 109140
QY 1201 agttgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1260
Db 109139 AGTTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 109080
QY 1261 gccccggggagaggttccagatcttcaatgactacagctgagcgagcagctatgac 1320
Db 109079 GCCCGGGGAGAGGGGTCCAGAACTTTCATGCTTACAGTGGGCGCTCCAGCTATGAC 109020
QY 1321 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1380
Db 109019 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 108960
QY 1381 catgccaagaaacccctcatatccaaaacccctgcaagatgacagcactacagcagca 1440
Db 108959 CATGCCAAGAAACCCCTCATATACCAAAACCTGCGCAAGTATACAGCTATAGGAGAA 108900
QY 1441 ggcacaggtactgacagcagcagcagcagcagcagcagcagcagcagcagcagc 1500
Db 108899 GGCACAGGCTACTGACAGCGGAGAGCGAGCGTCCGAGACCCAGAGAGTACTACAGACC 108840
QY 1501 ttccagcccaagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560
Db 108839 TTCCAGCCCAAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 108780
QY 1561 tccacagctgagcagctgagcagcagcagcagcagcagcagcagcagcagcagcagc 1620
Db 108779 TCCACAGCTGAGCGCTGATGCAAACTGGAAGAACTTTCCTTACAGCAGAGAGCGGCTC 108720

QY	1621	agacccggggcttccccgagggatacctgacacacagacacttcaatgcccctgtccta	1680
Db	108719	AGCACGGGGGCTTCGCCGAGGGATCACTGACACAGCACTTCATGCCCTGTCAAT	108660
QY	1681	ccctcccaacgagatgccacagctctgtgaacccagctgccaactcaagccctt	1740
Db	108659	CCCTCCCCAAAGGATGCCACAGCTGTGTGACACCCAGCGCTGGCACTCGAAGCCCTT	108600
QY	1741	cagaaggaacagctccctgagaacctgtgtcggaatctcagcctgcaagcctcagcgcg	1800
Db	108599	CAGAAGCAGACAGCTCCCTTGAGAACCTGTGTGGAATCTCAGCCTGCGAGCCTCAGCGG	108540
QY	1801	ctgacctacaggtggagaacatctccaaacagctccagcagctgtgctctccaaagct	1860
Db	108539	CTGACCTCACAGGTGGAGAACATCTCCAAACAGCTCCACGACGCTGCTCTCCAAGGCT	108480
QY	1861	gctgtgcgcagagaagaaggtgtcaagaacctcgtgtcaagagaccccaagagaacataa	1920
Db	108479	GCTGTGCGCAGAAAGAGTGTCAAGAACCTCGTGTCCAGGACCCCGAGACGACATAA	108420
QY	1921	agccagactcagcccccaaggaagcggtcactcagccgagcccgaggaacacacgctg	1980
Db	108419	AGCCAGACTGTGCCCCGGAAGGAGCGGCTACTCAGCGAGCCCGCAGGACACCGCTG	108360
QY	1981	tcagagccgcagcagcagcagccagctccacgaatcgcagagccgcagggagccgactac	2040
Db	108359	TCAGAGCCGCGAGCAGCAGCGCCACAGTCCACGCTCGAGCATCGAGCGCCGAGGCGGACTAC	108300
QY	2041	ctgagggctccgaggaaccactggagcgaggtctctactgtgaacccagggccgtggc	2100
Db	108299	CTGAGCGGCTCCGAGGACCCACTTGGAGCGGCACTTCCTACTGTCAACACGACCGCTGGC	108240
QY	2101	agccctgcagaggtcaacagcaactcgaagggccaaagcccgagtcggttcaactgtct	2160
Db	108239	AGCCCTGCCAGGCTCAACACGACTCGAAGCGCAAGCCGAGTCGCTGTCCACTGTCT	108180
QY	2161	gtgacctctctgacgaatgtccaccaaatctgacgactccttcagagactacacggc	2220
Db	108179	GTGACCTCTCCTGACGACATGTCCACAAATCTGACGACTCCTTCAGAGCCTACAGGC	108120
QY	2221	agtcgcgcctgcaagcttctccaaagttcgtggcggtgagggagctgtccggygtg	2280
Db	108119	AGTCTGCGCTCGACAGCTCTCCAAAGTTCGTGCGGGTAGCGGAGTGTCCGCGGCTG	108060
QY	2281	ctgctcagcgccctggacagagggacctggcctccgagatcctgggctcaggaagcc	2340
Db	108059	CTGCTAGCGCCCTGGCACHAGAGGACCTTGGCTCCGAGATCTCTGGGCTCCAGGAAGCC	108000
QY	2341	atcggtgaaagccgacaaagcttgggctgaagacccagctgtgtcaagacagcagc	2400
Db	107999	ATCGGTGAGAGGCCGACAAGCTTGGGCTGAAGCACCCAGCCTGTCAAGGACAGCAGC	107940
QY	2401	aagccacctctcgtggagaacacagcgcctgctggaactctgtggccaagagtgg	2460
Db	107939	AAGCCACCTCTCTGCTGGAGAACACACAGCGCTGCTGTGACTCTGTGGCCAAAGTGGG	107880
QY	2461	tggcccgccctggagagccgagggccctgcccgaactccttgagctggaacagggcgcg	2520
Db	107879	TGGCCCCGGCCTGGGAGCCGAGAGGCCCTTGCCTCCGAGTCTCTTGCAGCTGGACAAGGGGGG	107820
QY	2521	aatgccaaagactcagccagaggtgtttgaagaccccttcgtyggccttgcgtacgct	2580
Db	107819	AATGCCAAGGACTTCAGCCCAAGGCTGTTTGAAGACCTTCTGTGGCTTCTGCTACGCT	107760
QY	2581	gaccccaaaagacactgtctctctctctctctctgttggtaagacccacttggggttcct	2640
Db	107759	GACCCCCAAAAGACAACTGGTCTCTCTCTTGTGTACCAAGCCCAACCTTGGGGTTCCT	107700
QY	2641	gctcagaacccactacagagcttttgacttttccgacacacacacgctgccagctca	2700
Db	107699	GCTCCAGACCCCACTACAGAGCTTTTGTGACTTTTCCCGGACACAAACGCTGCCAGCTCA	107640
QY	2701	gcgagacgagccaaaccccttgcctggcagaggaagaaacctggggagatgtgttccaggg	2760

Dh 106559	GTGGGGAAGGCGCTCAACCAAGCTGCGCTCCAGCCGACGAAACCCGGCGCGCTGCTGTA	106550
QY 3841	gaccttcgacgacgacgcgcgaatggtgtctccaaagaccaagagacgaactcaaccgaagcgt	3900
Dh 106499	GGCTTCGACAGCAGCCGATGGGTCTCAAGACCAAGGAGACAACTCAACCGACGCGCT	106440
QY 3901	ggcaaggacacgagctccatgatctcttggtacgacacaaaccgaagagatcttccac	3960
Dh 106439	GGCAAGGACACGAGCGCTCATATCTTCCTCGGTCAACGACAAACCAAGGATATTTCCAC	106380
QY 3961	tccaaagcgcgagggcgctctcgagggcgcggtcccaacgtcggtgtccacaaagaatc	4020
Dh 106379	TCCAAAGCGGAGAGCGCTCTGAGGGCGGCTCCCAACTGCGTCCACCAAAAGCTC	106320
QY 4021	ctcgacacacgacacttgcgcgcgcacattcaaggtgtccagcgaccccgaaagagcg	4080
Dh 106319	CTCGACAAACAGCCACTGCGCGGCACATTCAGAGCTCTCAGCGCCCGCAAGAGAGGC	106260
QY 4081	aggtgtgacgacgagcgaggaaggttcccaaaccttggtgagcgacgaagctctcgacgg	4140
Dh 106259	AGGTTGAGCCACAGCGGACAAAGGTCCTCCCAACTGGTGAAGGACGACAAGTCCTGACCG	106200
QY 4141	cccccccatgcgtctaaagaagtggtgcgccttcacgtcgcggtgccacacgaagacgg	4200
Dh 106199	CCCCCTCATGCGCTCAAAAGAAATGCGGCTTCTTAGGGCGGCTCCCGCACAAAGCGG	106140
QY 4201	aaactgctcttgcgagacgagcagacgagagagacgaacgcgcagt-gcaatggygagat	4259
Dh 106139	AACCTGGCTTGGGAGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG	106080
QY 4260	gggaagagagagagcctgaaggttcccccaccccttcgaagagatgtctctccaaag	4319
Dh 106079	GGGAAGGAGGAGGCGCTGAGAGTTCCTCCGACCTCTTCAAGAGAGATGTCCTCCACAG	106020
QY 4320	aaagccaaagccacacaaagggacatgagagacgtgcgcacaaaggtccaccccgagagcc	4379
Dh 106019	AAAGCCAAAGCCCAACAAAGGCAATGCGACGCTGCCAATAAGCTCCACCCCGGAGACC	105960
QY 4380	ccc-gttctcgctctcagctctgcgtctc-gcacgccttcacaggggggcatafgaaccaag	4437
Dh 105959	CCGAGTGCCTGCGCACAAGTGCCTCGCGGCGACGCTTCAAGGGGCAATGAAGACAAAG	105900
QY 4438	gtgtgtgcaccccggaagggccggggtctgaagctggaagccatcgtgcagaagatcac	4497
Dh 105899	GTGTGTGCACCCCGGAGAGGGCGGGGCTTAACCTGGAAGCACTGCGAGAAGATCACCC	105840
QY 4498	tgcgcacacgtccaaagaagtgcacatgaagaagcgccagggcgctctctctgatatcttg	4557
Dh 105839	TGCGCCAAAGCTCAAGAAAGTGCATGTAAGAGCGCACAGGGGCTCTCTGTATATCTTGG	105780
QY 4558	agccacatcccttcgcagacaaagacggtgggtcctaaggtgtcttggtggcgaccagtgg	4617
Dh 105779	AGCCCATCTCTTTCGACAAACACCGTGGGCTCAAGGATGCTGGGGGCAAGCCAAATGGGG	105720
QY 4618	gttgaagaagggcctgtaaatgttgggcaacgggcagaagcttcccaactcttggtggctat	4677
Dh 105719	GTGGAAGAAAGGCTGTGAATAATGTGGGCAACGGGCAAACTCCCAATCTTGAGGCTGAT	105660
QY 4678	cogttatgcgaataatccacacacacgatctcttaaaagggaaactcagaacgtataagaa	4737
Dh 105659	CGATTATCAAAATCAACCAACCAACAGATCTTTAAAGGCAATCAAGAACAGTAAAGAA	105600
QY 4738	ctgtcttcctacgtactgttcaaaacacagagccttaacatcccgagggcgctcgagcgt	4797
Dh 105599	CTGTCTTCTTACTGACTGTTTCAAAACCGAGGCTTTCACATCCCGGAGGCGCTTCGAGCT	105540
QY 4798	gggggagactgcctctggcgctctaagaagaagagccggaagaaggccgggacagggccattga	4857
Dh 105539	GGGGGAGATGCGCTTGAGCGCTTAAGAAAGGAGCGCGAAAGCGCGGACAGGCGCCATAGA	105480
QY 4858	ctctccaaaggcccggtgtgaagaagggccatcttggcccggtctgctctcgactccg	4917
Dh 105479	CTCTCCAAAGGCGCGCTGGAAGAGGCGCTCATCTCTTGCGCCGCTGCTCTGACTCC	105420

[illegible]

QY 2461 tggcccgccgctgggagccggagcccgccgactccctgagctggacaaagggcgcc 2520
DB 11559 TGGCCCGCGCTGGGAGCCGGAGCCGCTGCGCGACTCTCTTGGAGCTGGACAGGGCGGC 11618
QY 2521 aatgccagacttaagccagggctgtttgaagacccttcgtggccttgactacgct 2580
DB 11619 AATGCCAAGGACTTACGCCAGAGCGCTGTGTTGAGACCCCTCCGTGGCCCTTCCTACGCGCT 11678
QY 2581 gaccccaaaagacaactggctctctctctctctgtttacaaagccacccctggggttcc 2640
DB 11679 GACCCCAAAAGACAACCTGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11738
QY 2641 gctccagaccccaactacagagcttttgactgtttccogagacacacccgctgcagctca 2700
DB 11739 GCTCCAGACCCCACTACAGAGCTTTTGAAGCTTTTCCGGAGACAAACCCCTGCCAGCTCA 11798
QY 2701 gggagacagcccaaccccttggctggcagaggaacacccctggggatcttggctccagg 2760
DB 11799 GGGAGACAGCCCAACCCCTTGGCTGGCCAGAGGAACCTTGGGGGATCTGTGTCACAG 11858
QY 2761 tggggattgacccctggcgagcttaccagggcctggagcagggtgggaagcctcagat 2820
DB 11859 TGGGGATTGACCCCTGGGAGCTTACCAAGGGCCTGGAGCAGGGTGGGAAGCCTCAGAT 11918
QY 2821 gcatcagcaaaagggacacccatagcttggcctcctggccttccagagaggac 2880
DB 11919 GGCATCAGCAAAAGGGACACCCATGAGGCTTGGGCTCTGGCTTCCAGGAGGAGAC 11978
QY 2881 ccccttgggagaggtggcctctgtccggggacttcaagcagagaggtgggtggg 2940
DB 11979 CCCCTTGGGAGAGGTGGCTCTGTGCCGGGGACTTCAAGCAGGAGGAGGTGGGTGG 12038
QY 2941 gtaagagagagcaggtgggctgctgagctggcccgaggtggccaaggtcacccgtgg 3000
DB 12039 GTAAAGAGGAGGAGCTGGGCTGCTGAGTGGCCCGAGGTGGCAAGGTGACCGGTGG 12098
QY 3001 ctgagagacagccgaactgttccacgcagacttggcagctccacactcga 3060
DB 12099 CTGAGGAGGAGCCGCACTGTCTTCCACCGCGACTTGGGGAGCTTCCACTGTGCA 12158
QY 3061 cccacagcaggaagagagcctggaagctgagagagtagtactcctctatgtagctc 3120
DB 12159 CCCACAGGAGGAGGAGGACTGGAGCTGAGAGGAGTACTCTCTCTATGTAGCTC 12218
QY 3121 ctggcagcccgagagagccctggcagcaggaacccgctgtccacagggcccactc 3180
DB 12219 CTGGCAGCCCGAGAGGAGGCTGGCATGACAGACCCGCTGTACCAAGGCCCACTC 12278
QY 3181 atctcaccagagagagagaggtgtgactcgaagccagccgctgggctctcg 3240
DB 12279 ATCTCACAGGAGGAGGTGAGAGGTGTGACTTCAAGGCCGCTGGGCTCTCCG 12338
QY 3241 tggccactctcagggaggtccgtcatcctctgctggccctacagtggcaccagtcgaag 3300
DB 12339 TGGCACTCTCAGGAGGAGTCCGTATCTCTCTGGGCCCTACAGTGGCAGTCAAG 12398
QY 3301 gtcagagctgtttgactcctctctgtccacatgaagccaggtgaagagggcctgat 3360
DB 12399 GTCCAGAGCTGTTTGTGCTCTCTCTGTCTACATTAAGCCAGGTGAAGAGGGGCTGAT 12458
QY 3361 gggagcgaactccagggattccacacactcgaagcctctctgcccagaagcccaac 3420
DB 12459 GGGAGCGAGCTCCAGGGGATTCACACACTTCGAGCGGCTCTCTGGCCCCAGAGGCCAAC 12518
QY 3421 aagcctgtgtgccagggcccatgcgaagagagcctgtgcagggggcgaagc 3480
DB 12519 AAGCTGTGTGCCCCAGGCGCCATGCGCAAGAGAGGCTGTGCCAGGGGCCAAGC 12578
QY 3481 ttacggagcctgggtgacccggggctgcccagggccagaggaactcccaatgcagggca 3540
DB 12579 TTACGGAGCCGTGGGTGACACCGGGGGCTGCCAGGCGAGGACTCCCTCCATGACAGGCA 12638
QY 3541 ccagtgctgcccagaagacacttctgctcactgacagagggcccccagggagac 3600

DB 12639 CCAGTGTCTGCCAAAGACTCTTGTCTCCTGAATCTGCACAGGGCCCCCAGGACAG 12698
QY 3601 atggaagggctggagcccgagccgggggctcgaaagggctcccgaggtgtgtact 3660
DB 12699 ATGGAAGGGCTGGAGCCCGAGCCCGGGGGGCTTCGGAAGGCTCCCCAGGATGTGTACT 12758
QY 3661 cgttctctcagggccctgagtgagcccgccagcccgccggagcccgagcctgacccacc 3720
DB 12759 CGTCTCTCAGGGCCTGAGTGAGCCCGGACGCGCGGAGCCCGAGGCTGACCCACC 12818
QY 3721 cctgaccccccagacaaactggggggaagcagcagcagcccttcaagtgggcaagcgg 3780
DB 12819 CTTTCACCCCGCAGAACTTGGGGGGCAAGCAGCAGCGCCCTTCAAGTCGGGCAAGCG 12878
QY 3781 gttgggaagccctcaccagaggtcctccagcccgccagcaacccggcgccctgctgtg 3840
DB 12879 GTGGGAAAGCCCTCACCAAGGCTGCTTCCAGCCCGCCAGCAACCCGGCGCCCTGCTGTG 12938
QY 3841 gctccagcagcagcccgatgggttccaaagcacaagagacagactcacccagcagcct 3900
DB 12939 GCCTCCGACAGCAGCCGATGGGCTCCAAGACCAAGAGAGACAGTCAACCCAGCAGCCT 12998
QY 3901 gccaagggcccgctcctatcttctgctcagcaccacaaacccagagatcttccac 3960
DB 12999 GGCAGGACCAAGCCTCATGATCTCTTGGTACGACCAACCAAAACCCAGGAGATCTTCCAC 13058
QY 3961 tccaaagcgggagggcctctgagggcggtcccaactgctgcccacaaagaagctc 4020
DB 13059 TCCAAAGCGGGAGGCTCTGAGGGCGGGCTCCCAACTGCGGTGCCACCAAGAAGCTC 13118
QY 4021 ctgcaaacagccacttggccgcacattcaagttctcagcagcccccagagaggagc 4080
DB 13119 CTGCAACACACCCATTTGCCCGGCACTTCAAGGTCTCCAGCAGCCGCCAAGAGAGGCG 13178
QY 4081 aggtgagccagcgggcaaggttccccaaactgttgcaggcagcaactctctgaccg 4140
DB 13179 AGGTTAGCCAGCGGGCAAGGGTCCCAAACTGTGTGAGGAGCAGCAAGCTCTCTGACC 13238
QY 4141 cccctcatggctcaaaaggaagtcggcgttcaatggcgccgtcccccacaaagaagcgg 4200
DB 13239 CCCCTCATGTGCTCAAAAGAAAGTGGGCTTCATGGCGCGGTCCGCCCAAGAGCGG 13298
QY 4201 aactgtgttggcagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4259
DB 13299 AACTGTGTTCGGGAGCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 13358
QY 4260 ggaagagagagagcctgaggggttccccccacttcaagagagatgtcttctcccaag 4319
DB 13359 GGAAGGAGAGAGCCTGAGGGTTCGCCACCTCTTCAAGAGATGTCTTCTTCCCAAG 13418
QY 4320 aaagcgaagcccaacaaaggggaatggagcctgccaagaactcccccccgagacc 4379
DB 13419 AAAGCCAAAGCCCAAGAGGCAATGGGAGCTGCCAAGCTCCACCCCGGAGACC 13478
QY 4380 ccc-attectgctcaagctcgcctctc-ggcagccttcagggggccatgaagacaaag 4437
DB 13479 CCGAGCTGCTGCTCAAGCTGCGCTCTCGGCGAGCCTTCAGGGGGCCATGAAGACCAAG 13538
QY 4438 gtcgtccaccccggaagggcgggcgctgaagctggaagccatcgtcagaagatcacc 4497
DB 13539 GTGCTGCCACCCCGAGAGGCGGGGGCTGAGGTGGAGCCATGTCGAGAGATCACC 13598
QY 4498 tgcggcgctcaagaaggttcgcatgtataagcgcaggggctcctcgtggaatcctg 4557
DB 13599 TCGCCAGCCTCAAGAAATTCGATGTAAGCGCAGGGGCTCTCTGTGTAATCTCTG 13658
QY 4558 agccatccttccgacaaagaccctgggtcaagggctgctggggcagccagtgagg 4617
DB 13659 AGCCCATCTCTTCCGCAAGACCGCTGGGCTCAAGGGTCTGGGGGAGCCAGTGGG 13718
QY 4618 gttgaagagagcctgtaaatgtggcccgccaggaactcccaactctggggctgat 4677

Db 13719 GTGGAAGAAAGGCTGTGTAATGTGGGACCGGAGCAGAAAGCTCCAACTTCTGGGGCTGAT 13778
 QY 4678 ccgttatcagaatcccaaccacagatctcttaagaagcaaatcatatgaagtaagaa 4737
 Db 13779 CCGTTATGAGAAATCCACCAACAGATCTTTAAAGGCAAACTCATGAAACAGTAAGAA 13838
 QY 4738 ctgtcttctactgaactgtttcaaaacccgaggtcttaacatccccggaagccctcaact 4797
 Db 13839 CTGTCTTCTACTGACTGTTTCAAAACCGAGGSCCTTCACATCCCGGAGGCTTCAGACT 13898
 QY 4798 ggggggagctgcgcctcgcccttaagaagaagcggaagcccgaggggagcc-atgg 4856
 Db 13899 GGGGGAHCTGCTTGGCCCTTAAGAAAGAGAGACGGGAAAGCGGCGAGGCGCCNATGG 13958
 QY 4857 actctcaaaagcccgctcgtgaagaagcgccctattctggccggctctgtctctgactcc 4916
 Db 13959 ACTCTCAAAAGGCGCGCTGAGAGAGGCGCCATCTTGGCCCGGCTCTGCTCCGACTTCC 14018
 QY 4917 ccgagagaagggccatgtgcaacaagaaggccagtgagaagaactctgtgtgaggaagca 4976
 Db 14019 CCGAGACAGGCGCGAGTGACACACAGAGGCGCAGTAAAGACACTCTGCTGAGAGAGCAA 14078
 QY 4977 gaagccaaagatgagagagctgtgacctgctctccagaccgccggaaggagggccctgca 5036
 Db 14079 GAAGCCAAAGATGAGAGAGCTGGGCGCTGGCCTCCAGGCCCCGAGAGGCGCCTGCA 14138
 QY 5037 gccccagaagaaggagcaagaagaagccagagccacaacaatacaagcagctatccaaagc 5096
 Db 14139 GCGCCAGACAAAGGCGACAGAAACAGCCAGGCGCACACCAACAAAGCACTATTCGACAGC 14198
 QY 5097 gaagcgctactactcgggcgcgaggccaagaagaacaactcttcaacctgtlaaggggctgc 5156
 Db 14199 GAGGCGCTCCTACTCGGAGGCGGCGCCAGAAACACCTCTTCACTTCACTTGAAGGCGCTGC 14258
 QY 5157 caagcgacagacagacaacagcagcgtcgcacctggatcccgagagccctgaatccgct 5216
 Db 14259 CAAGGAG 14318
 QY 5217 caagtacaattctctcttgaaagcgctgaggtcagaacgcgcggagcccgccctcttaac 5276
 Db 14319 CAAGTACATTTCTCTTGTGAGGCGCTGAGAGCTCAACAGCGGAGCGCCGCTTCTACCC 14378
 QY 5277 ctctgtcgaggtgagaaagcagagcgcgtttcaacaacatatgactgtgttaactccccc 5336
 Db 14379 CTTCTGTGGGAGTGAAGACGAGAGCGGTTCAACACATATGCACTGTGTCAATCCCC 14438
 QY 5337 tggagatgcgcccacagcccacaggaagccttctctctctctctctctctctctctctc 5396
 Db 14439 TGGAGATGCGCCCAAGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14498
 QY 5397 gtcctcgctctctcttgatgacagcggcggtccctccctgagcaactccctgagagctcat 5456
 Db 14499 GTCTCTGCTTCTCTTGTGAGAGCGCGGCGCTTCCCTGCGCACTCTCTGAGAGGCTCAT 14558
 QY 5457 cctgagagcgccgcccctctgcccctctctctctctctctctctctctctctctctctct 5516
 Db 14559 CCGGAGCGCGGCGCTCTCTTGTGCGCTCTCTCTCAAGAGAGAGAGAGAGAGAGAGAGAG 14618
 QY 5517 caagcgccctgagtaactcttgcctgttctgtcgtctctgccaacacccgggccaactcaa 5576
 Db 14619 CAAGGCGCTGAGTACTCTTGTGCTTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 14678
 QY 5577 ggaacttgggagacctctgtggcgcctactccctgaaacatgctctcccaaaagaagcc 5636
 Db 14679 GGAACCTTGGGAGACTCTGTGGGCGCTCTACTACTCTGAAACAGTCCCTCCCAAAAAGAAC 14738
 QY 5637 aaaaactcaaggaaggtgcggcgaagaagccacttgaaggaagccctcgatgagccctga 5696
 Db 14739 AAAACTCAAGGAG 14798
 QY 5697 gagaactcaagagtcagcaggtgtgagagctgcgcaactgcgggaagcccccag-- 5754
 Db 14799 GAGAAGACTCAAGAGTCCGAGTGTGACAGTGTGCGGCACTGCGGAGAGCCGCCAGGCGC 14858

QY 5755 ttagcgccagctgaccgcccagagcagggcccaactgagcaactgagccggcgctgc 5814
 Db 14859 TACGCGCCAGCTGACCCGCGCAAGAGGCGCCACTGCGCACAGTGTCCCGGGGCGCTGC 14918
 QY 5815 ccgagagctcagaagctgtactctgtgagtcggcgggaagatgagggcggaaggaagc 5874
 Db 14919 CCGAGAGCTGAGAGCTGCTACTGCTGTGATGAGCGGAGAGATGGGGCGAGAGAGCGAC 14978
 QY 5875 ccagcgcaagaaggtgcgaacatgagtcagcaagaagagctccgagagcccgagc 5934
 Db 14979 CCGAGCGCAAGAGGTGCGAAACAGTGTGACCAAGAGAGCTCCGACAGAGCCGAGG 15038
 QY 5935 gtaggcgcaagagagactggtgacatgagagctgtgcgtgtggaacggcgagctact 5994
 Db 15039 GAGGCGCCAGGA-CAGTGGGTCAAGAGCTTGTCCGTGTGAGACCGGCGCTTACT 15097
 QY 5995 ggtggcgagagagctcttggcgctgca 6022
 Db 15098 GGTGGCGGGAAGCTTTGGGCTCAG 15125

RESULT 4
 AB058723 5915 bp mRNA PRI 05-JUN-2001
 LOCUS Homo sapiens mRNA for KIAA1820 protein, partial cds.
 DEFINITION AB058723
 ACCESSION AB058723.1 GI:14017856
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens brain cDNA to mRNA, clone:h01321.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 (sites) Nakayama, K., Nakajima, D., Kikuno, R. and Ohara, O. Nagase, T., Prediction of the coding sequences of unidentified human genes. XX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
 JOURNAL MEDLINE 21245130
 REFERENCE 2 (bases 1 to 5915)
 Ohara, O., Nagase, T. and Kikuno, R. Direct Submission
 TITLE Submitted (27-MAR-2001) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: oohara@kazusa.or.jp)
 URL: http://www.kazusa.or.jp/muge, Tel: 81-438-32-3913, Fax: 81-438-52-3914

FEATURES
 location/Qualifiers
 source
 1..5915
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="h01321"
 /issue_type="brain"
 /note="vector: BluescriptII SK plus"
 458..5392
 /gene="KIAA1820"
 <458..5392
 /gene="KIAA1820"
 /note="Start codon is not identified."
 /codon_start=1
 /product="KIAA1820 protein"
 /protein_id="BAB41449.1"
 /db_xref="GI:14017857"
 /translation="PARTIAL PROTEIN SEQUENCE OF KIAA1820 PROTEIN. THE SEQUENCE IS BASED ON THE CDS OF KIAA1820 mRNA. THE SEQUENCE IS NOT IDENTIFIED."

CDS
 gene
 cds

[illegible]

QY	3349	gaggggacctgaatgagggagcgagctccaagggattccacacactcgsgcgcctctctggcc	3408
Db	3329	gaggggacctgattggagagcgagctccaagggattccacacactcgsgcgcctctctggcc	3388
QY	3409	caggaagcccaaaagacccctgcctctgcccagggcgccactcgcaaaaagagcgctctgcca	3468
Db	3389	caggaagcccaaaagacccctgcctctgcccagggcgccactcgcaaaaagagcgctctgcca	3448
QY	3469	cgggggcaaaagcttaagagacgttgagctcggttgcaacgggggctctccgagccgagatccc	3528
Db	3449	cgggggcaaaagcttaagagacgttgagctcggttgcaacgggggctctccgagccgagatccc	3508
QY	3529	ccatccgggcccaccagctctctcccaaaagacctctgtccctggaatctctgacaaaggccc	3588
Db	3509	ccatccgggcccaccagctctctcccaaaagacctctgtccctggaatctctgacaaaggccc	3568
QY	3589	ccccccaggaacagaatgaaaggggctctgagaccccaagccggggggctctggaaaggtcccc	3648
Db	3569	ccccccaggaacagaatgaaaggggctctgagaccccaagccggggggctctggaaaggtcccc	3628
QY	3649	aggatggtactctgttctctcaagggccctgagtgaagcccgagaccccgagaccccaaggtc	3708
Db	3629	aggatggtactctgttctctcaagggccctgagtgaagcccgagaccccgagacccccaaggtc	3688
QY	3709	ctgagacacacccctcgacaccccccagacaaactgsggggcaaggaacgagccgacctcaag	3768
Db	3689	ctgagacacacccctcgacaccccccagacaaactgsggggcaaggaacgagacgacctcaag	3748
QY	3769	tggggcaagcgggttgaggagacccctcaacccaagctctgctccagcccaagaacccggcc	3828
Db	3749	tggggcaagcgggttgaggagacccctcaacccaagctctgctccagcccaagaacccggcc	3808
QY	3829	ggccctgctctgagccctcgacacagaacccgaatgggtctccaaagaccaaaggaagacagctca	3888
Db	3809	ggccctgctctgagccctcgacacagaacccgaatgggtctccaaagaccaaaggaagacagctca	3868
QY	3889	cccagacagcctctgcaaaagacacagcgtctcatgatctcttggtcaacgacaaacccag	3948
Db	3869	cccagacagcctctgcaaaagacacagcgtctcatgatctcttggtcaacgacaaacccag	3928
QY	3949	ggaatcttccactcaagagcgsgggagccctctgagggccgggtctcccacactctgctggcc	4008
Db	3929	ggaatcttccactcaagagcgsgggagccctctgagggccgggtctcccacactctgctggcc	3988
QY	4009	accaaagactctctgcacacaacagcactctgcgcgcacatcaaggtctctcagcagcccc	4068
Db	3989	accaaagactctctgcacacaacagcactctgcgcgcacatcaaggtctctcagcagcccc	4048
QY	4069	cagaagggggcagaagttagacagggggaaggtctcccaaaactctgttcagaggaagaag	4128
Db	4049	cagaagggggcagaaggttagacagggggaaggtctcccaaaactctgttcagaggaagaag	4108
QY	4129	ctctctgagccggccctcactctgctgctcaaaagaaagtgcgacttcaatggcgcggtccccc	4188
Db	4109	ctctctgagccggccctcactctgctgctcaaaagaaagtgcgacttcaatggcgcggtccccc	4168
QY	4189	accaaagacggaactgtgtctcttgccgacagcgacgacgacgacgaacgacgt-gc	4248
Db	4169	accaaagacggaactgtgtctcttgccgacagcgacgacgacgacgaacgacgt-gc	4228
QY	4248	aacgggggagatgggaaggaaggaagggcctgagagtctcccaacccctcttaagaagatg	4308
Db	4228	aatgggggagatgggaaggaaggaagggcctgagagtctcccaacccctcttcaaaagaaatg	4288
QY	4308	tcttctcccaagaagccaagcccaagaagggaatvggagacgtctccacaagctccca	4368
Db	4288	tcttctcccaagaagccaagcccaagaagggaatvggagacgtctccacaagctccca	4348
QY	4368	cccccggaagacccc-attccgctcaagctgaccttc-ggacaccttcaagggggcc	4428
Db	4348	cccccggaagacccc-attccgctcaagctgaccttc-ggacaccttcaagggggcc	4408
QY	4428	cccccggaagacccc-attccgctcaagctgaccttc-ggacaccttcaagggggcc	4488
Db	4408	cccccggaagacccc-attccgctcaagctgaccttc-ggacaccttcaagggggcc	4468

QY 2950 gagcaggtgggtgctgcagtgcccgagtggtgcaaggtgacccggtggtggtgagagc 3009
Db 2320 GAGGAGGTGGGTGCTGCTGAGTCCCGAGGAGGTGGCAAGGTGACCGGTGGTGGAGGAC 2379
QY 3010 agcggcaactgctgttccacacgcgcgaacttggggacccctccactgtgccaaccacagc 3069
Db 2380 AGCGGGCACTGCTGTTCACCGCGGAGCTTCGGGGGACCTCCCACTGTGTGCCACCCAGC 2439
QY 3070 aggaagagagacacgaagctgagagagagtagtactccctctatgtagctctggtgagc 3129
Db 2440 AGAAGAGAGACCTGGAAGCTGAGAGAGAGTACTCTCTCTATGTAGAGTCTCTGGGAGC 2499
QY 3130 ccgagcagaggtggtgcatgacgagacccgcgtgaccccaaggcccccactcatatgcacc 3189
Db 2500 CCCGAGCAGAGGCTGGCATGCGAGGACCCGCTGTACCCCAAGGCCCACTCATCTGCAAC 2559
QY 3190 aagagagagtgagaggtgctgtagtactcaagcgcggtggtgctctccgtgcaactc 3249
Db 2560 AAGAGAGAGTGGAGAGGTGCTGGATCTCCAGGCGGCTGGGGGCTCTCCGTGCCACCTC 2619
QY 3250 tcagggagtcctgcatcctgctggtggtcctacagtggtggcaccagtgcaaggtccagc 3309
Db 2620 TCAGGGAGTCCGTCTCATCTGCTGGGCCCTACAGTGGCGACCGAGTCAAGGTCCAGAGC 2679
QY 3310 tggtttagtctctctgtccacatgaagcagtggaagggggtggtggtggtggtggtggtg 3369
Db 2680 TGGTTTGGAGTCTCTCTGTACACATGAAGCCAGGTGAAGAGGGGCTGATGGGGAGCGA 2739
QY 3370 gtcacagggattccacactcctgagcgtcctctggtggtggtggtggtggtggtggtggt 3429
Db 2740 GCTCAGGGGNTTCCACCTCTCGAGCGCTCTCTGGCCGAGAGCCCAACAGCCCTGCT 2799
QY 3430 gtgcccagggggccatcgcaaaagaagagctgtgcaaggggggcaaaagcttaccgagc 3489
Db 2800 GTGCCGAGGCGCCATCGCAAGAAAGAGCCGTGTGCCACGGGGCAAAAGCTTACGGAGC 2859
QY 3490 cgtcgggtgacccgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 3549
Db 2860 GGTGGGTGTCACCGGGGGTGTCCGAGCGCGAGGACTCTCCATGCAAGGCGACCAAGTGTG 2919
QY 3550 cccaagaccttctgctcctgaaatctgcaaggggggggggggggggggggggggggggggg 3609
Db 2920 CCACAAAGACCTCTTCTCTGTAATCTGTCACAGGGGCCCCCCCCAGGGGACAGATGAAGG 2979
QY 3610 gctgagcccccaggg 3669
Db 2980 GCTGAGGCCGCCGCGGGGGGCTCTGGAAGGGGCTCCCGAGGATGTACTGTCTCTC 3039
QY 3670 acggtcctgagtgagcccgacgcccgcgaccccgagcctgacaccccccctgcaccc 3729
Db 3040 AGGGCCCTGAGTGAGCCCGCGACGGCGGGACCCCGAGGCTTGACCCACCCCTGCAACC 3099
QY 3730 ccagacaaactgggggggcaagcagcagcgccttcagtgctggggcagcgggtgggaag 3789
Db 3100 CCAGACAACTGGGGGGCAAGCAGCAGCGCGCTTCAAGTGGGGCAGCGGTGGGAAG 3159
QY 3790 cctcaacccaaaggtcctccagcccgacaccccgccggtggtggtggtggtggtggtggt 3849
Db 3160 CCTCACCAAGGCTGCTTCAGCGCCGACAGCCCGGGGCTGCTGCTGCTGCTGCTGCTGCT 3219
QY 3850 agcagcccgatgggttccaaagcaagagagacagactcacccagcagcctggtgcaaggac 3909
Db 3220 AGCAGCCCGATGGGCTCGAAGACCAAGAGAGAGAGTCACTCCAGCAGCGCTGGCAAGGAC 3279
QY 3910 cagcgtccatgatacttctggtgacgacccaaacccggagatcttccactccaaaggg 3969
Db 3280 CAGCGCTCATGATCTTGGTGTGACGACCAAAACCAAGAGATCTTCCACTCCAAAGCGG 3339
QY 3970 cggagccctctgagggcgctcccaactcctgacacccagagctcctccagac 4029
Db 3340 CGAGGGCCCTGTGAGGGCGGCTCCCTCACTGCGCTGCGACCAAGAGTCTCTGACAC 3399
QY 4030 agccacttgcgcgcacattcaaggtctccagcagcccccaagaggggaggtgagc 4089

Db 3400 ACCCACTTGGCCGCCACATTCAAGGTCTCCAGACGCCCCAGAGAGGGGAGGGTGGC 3459
QY 4090 cagcgggcaaggggtcccaaaactggtgagcagcagcaagctctctgtgacccggccctccat 4149
Db 3460 CAGCGGGCAAGGGTCCCAAAAGCTGTGAGGAGCAAGAGCTCTGACCGGGCCCTCCAT 3519
QY 4150 ggcctcacaaggaagtcggccttcagtcgagcgcgtcccccacaaagagcgaactgctc 4209
Db 3520 GGGCTCAAAAGGAAGTCTGGCTTCTATGGCGCGGTCCCCACCAAGAGCGGAACCTGTC 3579
QY 4210 ttggtgcacggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4268
Db 3580 TTGCGAGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3639
QY 4269 gtaggctcagaggttcccccacactcttcaagagatgtcttctccaaagaccaaag 4328
Db 3640 GAGAGGCTGAGGGTTCCTCCCACTCTTCAAGAGGATGTCTTCTCCCAAGAAAGCAAG 3699
QY 4329 cccacaagggcaatggcagcgtgcaaaagctcccccacagcagcagcagcagcagcagcagc 4387
Db 3700 CCCACCAAGGGCAATGGCGAGCCTGCCACAAAGCTCCCAACCCCGGAGACCCCGATGCC 3759
QY 4388 tgcctcaagctgcctctc-ggcagccttccaggggggcatgaagaccaaaggtgctgca 4446
Db 3760 TGCTCAAGCTCGCTCTCGGGCAGCTTCCAGGGGGCATGAGACCAAGGTGTGCCA 3819
QY 4447 ccccggaagggcggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 4506
Db 3820 CCCCGAAGGGCGGGGCTGNAAGTGAAGCCATCTGTGAGAGATACCTTCGCCCAGC 3879
QY 4507 ctcaagagttgcctgtaagggcgagggggtggtggtggtggtggtggtggtggtggtggtggt 4566
Db 3880 CTCAGAAGTTCATGTAAAGCGCCAGGGGCTCTCTGTAATCTCTGAGCCCATCC 3939
QY 4567 ctttccacaagacgtggtcaggggtggtggtggtggtggtggtggtggtggtggtggtggt 4626
Db 3940 CTTTCCGACAAAGACCGTGGGCTCAAGGGGTCTGGGGGCGAGCCCGAGTGGGGTGAAGAA 3999
QY 4627 ggcctggttaagtggggcagggcgagagctcccaacttctggggtggtggtggtggtggt 4686
Db 4000 GGCCTGTGTAATGTGGCGACCGGGCAGAGCTCCCAACTTCTGGGGCTGATCCGTTATGC 4059
QY 4687 agaatccacacacagatccttaaaaggcaactcatgaacagtaagaactgcttct 4746
Db 4060 AGAAATCCAAACCAACAGATCTTAAAGGCAAACTCATGA----- 4099
QY 4747 actgactgtttcaaaacccgagccttcacatcccgagggcctgagcctggtgggggact 4806
Db 4100 ----- 4099
QY 4807 ggcctggcgtcagaagagagagcgggaaaggcgagggggggggggggggggggggggggg 4866
Db 4100 ----- 4099
QY 4867 ggcctggtgagagagggccctctcttggcccggtctgctcctgactcccgagacagg 4926
Db 4100 -----ACAGG 4104
QY 4827 ggcctggtgagagagggggcgagtgaggaactctggtgagggagcgaagaccaaag 4986
Db 4105 GCGAGTGGCACACAGAGGGCGAGTGAAGCAACTCTGTGTGAGGAGGAGGAGGAGGAGGAG 4164
QY 4987 atggagagctggggtcctccacagcccccgaggggagggcctgagcagccccagaca 5046
Db 4165 ATGGAGAGGTGGGCTGCTCCAGCCCCCGAGGGGCGAGGCGCTGCCAGCCCGCAGACA 4224
QY 5047 agggcacagaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5106
Db 4225 AGGGCACAGAAACAGCAGCGCCACCAACTACAGAGCTATTCCAAAGCGGAGGCGGCTC 4284
QY 5107 actcggcggggggcaagaaacacactcttccactgtaaggggggtgcaagcgagcga 5166

Db 4285 ACTGGGGCCGGGCAAGAAACACACCTCTTCAACCTGTAAAGAGGAGGAGGCAAGCA 4344

QY 5167 cgaacaaacagatgctgccccgagatcccgagagagccttaactcgcctaaatatt 5226

Db 4345 CGACAGACAGAGTGTGCCCCGAGATCCCGAAGAGCCTTAATCCGCTCAAGTAACTT 4404

QY 5227 tccctctgcaagcagctgagtgacagacagcagcagcagcagcagcagcagcagcagc 5286

Db 4405 TCCTCTTGCAGAGGCTGAGGTAGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4464

QY 5287 gtgagaaacagagacagcttaccacacacacacacacacacacacacacacacacacac 5346

Db 4465 GTGAGAGAGCAGAGAGAGGCTTACACACACACACACACACACACACACACACACAC 4524

QY 5347 ccaagagcccaagagagagccttccctcctcctcctcctcctcctcctcctcctcct 5406

Db 4525 CCAAGAGCCCAAGAGAGAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4584

QY 5407 tccctgagatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5466

Db 4585 TCCTTGAATGACAGCAGGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4644

QY 5467 cagccctctctgccccct 5526

Db 4645 CGGCGCT 4704

QY 5527 agtactctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 5586

Db 4705 AGTACT 4764

QY 5587 gacctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 5646

Db 4765 GACCTCTGCTGCT 4824

QY 5647 gagaaggtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 5706

Db 4825 GAGAAGGTGCTGCT 4884

QY 5707 aaaggtccagatgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 5764

Db 4885 AAAAGGTCCGAGTGTGCT 4944

QY 5765 gctgacccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5824

Db 4945 GCTGACCCGAGCAGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5004

QY 5825 cagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 5884

Db 5005 CAGAGCTGCTGCT 5064

QY 5885 aagggctgcaaacatgagtgacagagagagcgtccgagagcagcagcagcagcag 5944

Db 5065 AAAGG----- 5069

QY 5945 gagcactgagtgatgagagcgtgctgctgctgctgctgctgctgctgctgctgctg 6004

Db 5070 --ACACTGAGTGCATGAGAGCCTTGTGCGGTGTGAGACCGAGCGCTCTACTCTGT 5127

QY 6005 aaagctcttggctgag 6022

Db 5128 AAGCTCTTGGGCTGAG 5145

RESULT 6

LOCUS HSA271790 5667 bp mRNA PRI 17-JUN-2001

DEFINITION Homo sapiens mRNA for retinoid-acid induced protein 1 (Rai1 gene).

ACCESSION AJ271790

VERSION AJ271790.1 GI:12053792

KEYWORDS Rai1 gene; retinoid-acid induced protein 1.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 5667)

AUTHORS Seranski P., Hoff C., Radelof U., Hennig S., Reinhardt R., Schwartz C.E., Heiss N.S. and Poustka A.

TITLE Rai1 is a novel polyglutamine encoding gene that is deleted in Smith-Magenis syndrome patients

JOURNAL MEDLINE

REFERENCE 2 (bases 1 to 5667)

AUTHORS Seranski P.

TITLE Direct Substitution

JOURNAL Submitted (31-JAN-2000) Seranski P., Molecular Genome Analysis, Deutsches Krebsforschungszentrum Heidelberg, Im Neuenheimer Feld 280, 69120 Heidelberg, GERMANY

FEATURES

SOURCE

1. 5667

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="17"

/map="17p11.2"

1. 5589

/gene="Rai1"

1. 5589

/function="unknown"

/product="retinoid-acid induced protein 1"

/codon_start=1

/translation="MOSFRRCGPHGKQNTYQSEFRLNTROPSQAGLSQDROR

LLAKIYINPQYPSYBEGAGTSSGAAVAADYHHGSAALPTQOGLQGRAPRGV

QDSPPYRIVAGESLQAWGAPPP

GAVPFRKSLHYQPP

APTAQPHRPLTASSSLAGQVOMLHAYQSGRLSYDQCCQCCQCCQCCQCCQCCQCCQCCQ

AAETLHYONLAKTOHYGQGGQYQDPAAYTPROYTOYFSSSSSSHPASVGRSPSY

SNRPSLIMPENPYSQDPLSTGAFFAGTIDSHHMLPNSPDPATSSPADITAT

SOYENISNTVOQLLSKAAIPQKVKYKUTSRIPROKQCHSPGSGYSAERAGPL

SEPSSTPOSTHAFORADTLSSDPLENRFQKQAGSPARVNSKAKESYST

CSVTSEDMSTKSDSFQSLHSDPLDSFQFVAGRDQPLLALADLSELGL

LOALIERADKAMAEALSLKSDSPPLSLHNSACIDVAKASAVRGPPEALPSSL

OLDKGNANQDSPLFEDPVAFAAPPDKKTGTGLSFGRPTGAPADPTAFCF

PDRTAASSADSNFPAPEENLGDACPMWGHQBELTGLBGGKASDGLSKDTHEA

SATGROEDPDGRKVASLIDPKQKQEVGKYEAGGLIGLQEPYAKADRLDTRKCC

SLDRDDELPLPTSKNEDLEAEYSSYCLILSPQRPQMDPLSKAPDLICTKEA

VEEVILDSKAGWSPCLHSESVYLLGPTVGVESVQSWESSLSHMKPGEEDGREA

PGDSTLSDSLAOKPNPVAPEALPAKPEPRKSLRERVHRELLEADSPCRAPV

LPKOLLPECTGPPQGMGADAPGASBGLPRMCTSLALSEPRTPGGLTTP

PAPKDLGKORAAFKSGKRVGKSPKASPSNPALPVASDSSPGSKTKETSDPS

TPKQDORSWILASRTKQOETLHSHKRRSPRRRLPDRATKTLIDNSHLPMKRVSSP

QKGRYSORARVPKAGSKLDRPLHLAKKSKANAPVPTKRLVTRSSSSSNA

SGNGVGRKEPESPTLFRMSPEKAKPKKNGEPARKLPPEPTDACLKLASBA

FOGAKMTKLPKRGRLKLEAYOKITSPILKFKOKAKGASPNPGLSLDNDRG

LKAGGSPVYVEGLVWVGQKLPFGADPLCLNPRLSLGKLMNAGTGGASED

NSGGGKKRKMELGLASQPEGRPOQTOAKOGHNYSYKRRRLRGRAKNT

TSSPCGKAKRROOYPLDPAPEPRITKLTISCKRISDRSDRPPASPVYREKDA

PLTTCYVNSPDAPKHKRTSSSSSSSSSSSSSDAAGASLALTRGSLPRESL

PLSTVHLSPVYSKALSTQVCCQNPANFKLDLDGPTTPHCLPKKRLKLEK

VREBGEASLPLERTKGPDECAAAATAGAPDPDPKOPPLKASRLSR

QSCYCCDGEDGEGEAPADKGMVHECAVWTGVLVAGLFLQELAKVAVDNG

GYVALAPGPESLPGIRFASLFOGSPWCCPYLALSPSSMSGPAKAKGPGLAGAT

KDASPTTTRVPAOVARRGCTGNKSPKSGKSLTLPKSSSNPSPISLSPPLIQ

OFFPSPVCLDCAWAPGTWK"

BASE COUNT 1181 a 1962 c 1664 g 860 t

ORIGIN

Query Match 76.8%; Score 4627.6; DB 9; Length 5667;

Best Local Similarity 92.7%; Freq. No. 0;

Matches 511; Conservative 0; Mismatches 9; Indels 396; Gaps 10;

QY 490 atgcagctcttcgagaaggtgtgttcacatgagaacacagactaccagagacc 549

Db 1 ATGCAGCTCTTTTCAGAAAGGTGTGGTTTCCATGGGCAAAACAGAACTACACAGCAGACC 60
Qy 550 tgcaggaatacatcagcctcagagaaattacagagcagccagagtcagggccgagtcagagtcgc 609
Db 61 TCGCAGGAAACATCAAGCTTAGAGAAATACAGGACAGCCGAGTACAGGCGGGCTAAGCTGC 120
Qy 610 gacggcgagcggtgtgtccgaagagctattataaaccagcagccttaccagagctatag 669
Db 121 GACCGCAGCGGCTGTTCGCAAGGACTATATTAACCCAGCCTTACCGGAGCTATGAG 180
Qy 670 gctgcgctgacagccctctgacactcagcgcagcgcggtgtgcgcgcagcagtaaccacga 729
Db 181 GGTGGGCTTGGCAGCGCCTCTGGCACTGCACGCCCGGTGGCCGCGGACAGATACACCGA 240
Qy 730 ggcagcaagccctgcacacagcaagcctgcagggagcgcgttctccctgcttac 789
Db 241 GGCAGCAGGCGCTGCCACACAGCAAGCGCTGACGGGAGCGCGCTTTCCTGGCTAC 300
Qy 790 ggcgtccaggaacagcagccctaccagcgccgctatgtgtgagggagagccttaagct 849
Db 301 GCGCTCCAGGACAGCAGCCCTTACCAGCGCGCTATGCTGTGTGAGGAGCGCTTACAGGT 360
Qy 850 tggggggccacacagccacacacacacagcgcgcagcactacctgcaggggtggcgaag 909
Db 361 TGGGGGCCCCACAGCCACACCCCCACAGCCCGGAGCCACTACCTGCAGGGGTGGCCAG 420
Qy 910 tatgatgaactgatgaaaaagacagcagtcgccccagcagcagcagcagtcagagcag 969
Db 421 TATGATGAGAACTTGATGAAAGACAGCAGTGCCTGCCCGCCAGCAGCAGTATGAGAGCAG 480
Qy 970 ggcgcaggtgcctcttgcactcctcctcagcctcagcctcagcagcagcagcagcagcag 1029
Db 481 GCGGCCAGGTGCCCTTTCGGACTCATCCTCTGCACCTGCACGACGCCACCGCCGCCAG 540
Qy 1030 cagccctggcatacccaagctccaaagcagagcagtcgcaagcagcagcagcagcagcagc 1089
Db 541 CAGCCCTGGCATACCCAGCTGCAAGGAGCAGAGCTGCAGAGCAGCATTTGCCCTCCCT 600
Qy 1090 ctgccttcccaggggtacccattctcagcattcagcattcagcattcagcattcagcattc 1149
Db 601 CTGCCCTTCCCGCAG----- 615
Qy 1150 acctactcctctgtccaggggtgtggcagggggccactcctataagagttgaca 1209
Db 616 -----GGGTGTGGGCGAGGGG--CCACTCTTATAAGAGTTGCACA 654
Qy 1210 gcacgactgcccagcccatgacagcgctgactgcccagctccagcctgcccgcggg 1269
Db 655 GCACCGACTGCCACGCCCATGACAGCGCGCTGACTGCCAGCTCCAGCCTGGCCCGGGG 714
Qy 1270 cagcgggtccagaattcttcatgctaccagtcgggcgcgcctcagctatgaccagcagcag 1329
Db 715 CAGCGGTCCAGAACTTTCATGCTTACCAGTCCGGGCGCGCTCAGCTATGACACAGCAG 774
Qy 1330 cag 1389
Db 775 CAG 834
Qy 1390 gaaacctccattacaaaacctgcgaagtatcagcactacgggcagcagcagcagcagcagc 1449
Db 835 GAAACCTTCATTACAAACCTTGGCCAGTATCAGCAGTACGGGAGCAGAGCCGAGGGC 894
Qy 1450 tactgccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1509
Db 895 TACTGCCAGCGCGCAGCGCTCCGGAACCCAGAGCAGTACTACAGACCTTACGCCCC 954
Qy 1510 agctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1569
Db 955 AGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1014
Qy 1570 tgcgcgctgagcgaacctggagaacttccctcagcagcagcagcagcagcagcagcagcag 1629
Db 1015 TCGCGCTGTATGCCAAACCTTGGAACCTTTCCCTACAGCCAGCAGCGCGCTCAGACCGGG 1074

Qy 1630 gccttccggcgaggatcaatgacacagcagcagcagcagcagcagcagcagcagcagcagc 1689
Db 1075 GCCTTCCCGCAGGAGTCACTGACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1134
Qy 1690 acgagatccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1749
Db 1135 ACGGATGCCACAGCTCT----- 1152
-Qy 1750 aagtcctcagagaaactcgtcgtcagcagcagcagcagcagcagcagcagcagcagcagc 1809
Db 1153 -----CCTGCAGA-CCTCAGCGCTGACCTCA 1179
-Qy 1810 cagtgagaaatctccacacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1869
Db 1180 CAGGTGAGAACTATCCACACACCGTCCAGCAGCTGTCTCTCCAGGCTGTGTGCG 1239
-Qy 1870 cagaagaaggtgtcaagaacctgtccagagcagcagcagcagcagcagcagcagcagcagc 1929
Db 1240 CAGAAGAAAGTGTCAAGAACTGTCTGTCAGGACCCAGCAGCAGCAGCAGCAGCAGCAGC 1299
Qy 1930 tgcagcccgaaaggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1989
Db 1300 TCGAGCCCGAAGGAGCGGCTACTCAGCGAGCGCGCAGCAGCAGCAGCAGCAGCAGCAGC 1359
Qy 1990 ccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2049
Db 1360 CCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1419
Qy 2050 tccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2109
Db 1420 TCGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1479
Qy 2110 aggttcacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2169
Db 1480 AGGCTCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1539
-Qy 2170 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2229
Db 1540 CCGTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1599
-Qy 2230 ctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2289
Db 1600 CTCGACAGCTTCTCCAACTGCTGGCGGTGAGCGGAGCTGTCGGCGGTGCTGCTCAGC 1659
-Qy 2290 gccttcgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2349
Db 1660 GCGCTGGCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1719
Qy 2350 aagccgacaaagcttggcctgaagcagcagcagcagcagcagcagcagcagcagcagcagc 2409
Db 1720 AAGCGCGAAGAGCTTGGGCTGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1779
Qy 2410 tctcgtcagagacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2469
Db 1780 TCTCGCTGGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1839
-Qy 2470 cctggggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2529
Db 1840 CQTGGGAGCAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1899
-Qy 2530 gacttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2589
Db 1900 GACTTCAGCCAGGGCTGTTTGAAGACCTTCCTGGGCTTGGCTAGCAGCAGCAGCAGCAGC 1959
-Qy 2590 aagaacactggtcctctctcttcttgggtacaaagccaccccttgggttctcgtcagcagc 2649
Db 1960 AAGACAACTGCT 2019
-Qy 2650 cccatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2709
Db 2020 CCCACTACAGCAGCTTTTACTGCTTTTCCGCGAGCAACAGCAGCAGCAGCAGCAGCAGCAGC 2079

[illegible][illegible]

[illegible]

polya.signal
polya.signal
polya.site
BASE COUNT 1612 a 2335 c 2027 g 1248 t
ORIGIN

Query Match 60.8% Score 3664; DB 10; Length 7222;
Best Local Similarity 80.6% Pred. No. 0; Mismatches 935; Indels 153; Gaps 14;
Matches 4518; Conservative 0;

QY 469 cacagataacagcccgagctatgaagctcttcgagaaggtgtgtttcaatgcaaa 528
DB 288 CCGAATATACAGCCCGAGTCTATGAGTCTTTTCAGGAAGGTGTGTTCCATGGCAA 347
QY 529 caacagaactacagagaccccgagaaacatcacgccttagaattacaggaacgcg 588
DB 348 CAGCAAAACCTACCCAGACCTCCAGAGACATCCGCGCTGCGAAGACTCAGAGAGCGG 407
QY 589 agtcagcgcggttaagctgtgacagcgagcggtctctgcgaagagcttatataccg 648
DB 408 GATCAGCTGAGCTAAGCTGTATGAGGAGGCTCTGAGCAAGAGCTCTACAGCCTT 467
QY 649 cagccttaccccgagctatgaggtgtgctgtgacgcccctctgacatgtgacggtgtg 708
DB 468 CAACCTTATACAGGCTATGAGGCTGAGGCTGATACCTTTCGAGAGGAGGCTGCAAGAGA 527
QY 709 gccgcgagcaagtaacacaggaagcgacgaagccctgcacacagcaagcgctgcaggg 768
DB 528 GCTGAGAGCAAGTACACAGGAGGAGCA-----AATCCTGCGAGGGG 569
QY 769 aggcgcggtcttcctggtgagcggttcagagagagccctacagagcgctatgtct 828
DB 570 AGGCAAGCTTTCAGAGCTA---TCTTACAGAGAGAGCCCTTACAGAGGAGCTTCTCT 626
QY 829 ggtgagagagccttcaggtctggggggccacacagccacccacacagcgcgagcca 888
DB 627 GGGAGAGAGGTCTTTCAGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 686
QY 889 ctactgcaaggtgtgccaagtatgatagaacttgaataaagaagaagtg---ccc 945
DB 687 CTGCTGGGAGAGTGAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 746
QY 946 ccagagagagatgtgagagagaggggagagaggttccttcgagactcaactcctgcac 1005
DB 747 CCAAGAGAGAGTACCTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 806
QY 1006 gtccagagcagccgcccagagagagccctgtgcatccccaagctccacaagagagaa 1065
DB 807 GTC---CCAGCAACAGAGGCTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 863
QY 1066 ctgagagagagatgtcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1125
DB 864 CCAAG 923
QY 1126 tcccaactcctcccaactcctcccaactcctcctcctcctcctcctcctcctcctcctcct 1185
DB 924 TCCAGAGTCTTCTCCTACCTCCTCCTACCTCCTCCTACCTCCTCCTACCTCCTCCT 983
QY 1186 gcccaactcctataagattgacagcagcagctgcgcagcccatgacagagcgctgact 1245

SSSRIPDCAATKLLPANHLPTAFVSSGFOKEKMSORVYKPGTGKLSDRPLHT
LKRKAPAPVPAKRSILSNNGSGDREERERSSSGLLRMAAPORAPRSGSE
PPPPPLEPPACMGISTOSSISPSAVYTKYLPKRRKRLKALVQITISPOLKTLAC
KVAGPPEPTSPALPERPGSPAGABEGLGDMPLASGADPLCRPSRSLKAL
KLNSKRLSSADCPRAKAFMSPEPTSLTATAPAKSRGRGTGLPKRPLKRP
CPGQPLIALPHRASSTGGEDNSGSGKPKPTHEELPAQSPBPPOPLAKOK
PGQASYSKRRRLSRGKGTAAHSPCKRATRRROOVPLDPAPEIRKLYISC
KRLRDSSTPAEPYPRKRDATYMYCYVNHGAPVYSKALSTGLVCCLCQNPANFDLG
EPAGSLTTPGGSVLQORPSLPSSTMSHGLPYKALSTGLVCCLCQNPANFDLG
DLGQYVPEHCLPRKPKLEKARLEGLLEASLPLERLGLDLSSTIAPATAT
ITTPALGRLSRPDGPADPAKQGPETLRSRLQSGCCDQDGGGEVEVAQDS
RKHECKEAPTEPGEPTQEHWHYHBAKWTSGVYLVAGKRLQEMAKYAVDMVNG"

DB 984 GCCCACTCTACAGAGCTGCAAGACACCATCTGCCCCACCTCATGATAGGCCGATGAGT 1043
QY 1246 gccaagtcacagcctgtgcccgggggagcggttcgaagaatcttcattgctacagctgtggc 1305
DB 1044 GCCAATGCAACCTGCTTCCAGAGGAGAGGCTCCAGAACTTCAAGCTTCAAGCTTGCC 1103
QY 1306 cgcctcaactatgacagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1365
DB 1104 GCGCTTGCTGACG-----AGCAAGAGAGAGAGAG 1133
QY 1366 ctccagagcgggacacatgcccagaagaaacctccatccaaacactccgccaatgaag 1425
DB 1134 CTTCAGAGGCCGTCACACACACCCAGAGAACTCCACATCCAGAACCTCGCCAGATGACAA 1193
QY 1426 caactaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1485
DB 1194 CCAATGAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1253
QY 1485 cagtaactac 1545
DB 1254 CAGTATATACCAACCTTTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1313
QY 1546 tcaactctcactacagtcac 1605
DB 1314 TCCCTCTCTATAGCTCCACCCGCTCACACATGATGCGCAATCTGAGAACTTCCCTAT 1373
QY 1606 agccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1665
DB 1374 AGCCAGACCCCTTATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1433
QY 1666 atgcacatcctcactccctccacacagcagcagcagcagcagcagcagcagcagcagc 1725
DB 1434 ATGCCCCCTGCTTAAACCGTCCCAACAGATGCTGCTGAGGAGGAGGAGGAGGAGGAG 1493
QY 1725 aactgcaagcccttcagaagaaagctccctcagagacacacacacacacacacacacac 1785
DB 1494 AACTGCAAGCCCTTCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1553
QY 1785 cagagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1845
DB 1554 CAGAGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1613
QY 1845 ctgctctcagaagcgt 1905
DB 1614 TTGCTGTCCAAAGCTTACCATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1673
QY 1905 ccagagcagcatataaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1965
DB 1674 CCAAG 1733
QY 1965 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2025
DB 1734 GCGGGGACACCGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1793
QY 2025 caggagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2085
DB 1794 CAG 1853
QY 2085 aacagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2145
DB 1854 AGCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1913
QY 2145 gtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2205
DB 1914 GGTGCACTGCTGT 1973
QY 2205 cagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2265
DB 1974 CAG 2033
QY 2265 gactgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2325
DB 2034 GACTGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2093

QY 2326 gggctgaggagccatcggtgagagggcgacaaagcttggtgctgaagcaccagcctg 2385
DB 2094 GGACTGAGGAGGACCATGTTGAGAGGCTGACAAAGGCTGGGCTGAGGCTTCCAGCCTG 2153
QY 2386 gtcaaggagcagcagcagccactctctgctgagaaacacagcgcgcctgactct 2445
DB 2154 OCCAAGGACAAATGCAAGCAACCTTCTCACTGAGAAACATGCGCTGCTGGACACT 2213
QY 2446 gtggcagaagtgctggcccgccgctggggagccgagagccctgcccactcttgag 2505
DB 2214 GTAGCAAGAGCTTCATGGTCAAGCCAGGGAGCCAGAGACCTCCCTGAGCCTTTGAG 2273
QY 2506 ctggcaagggcgcaatgcgaagcacttcaagccggctgtttgagaccctctcctg 2565
DB 2274 CTGGACAGGCTGACAGCACCACAGGACTTCAGCCCGGGGCTGTTGAGAGCCCTCTG 2333
QY 2566 gcocttgtaacyctgacccccaaaaaagacaaactggtctctctctctctgttatacgaagcc 2625
DB 2334 GCGTTTGGCCACACTGACCCGAGAAACAAAGCAGTCCCTGTCTCTCGGCACCAAGGCT 2393
QY 2626 acccttggttctctgctcaacacccccactacagcagcttttgacttttcccgacaca 2685
DB 2394 TTGCTAGGAGACTGCCACTCTGACCCACCCACAGCAGCAITTGACTGCTTCCAGATACA 2453
QY 2686 accgctgcagctcagcgacagcgcaccccttggctggcgagggaaaacctgggg 2745
DB 2454 CCCACTGCCAGCTCGGTGGATGGTGCACAAACCCCTTGGCTGGCCAGAGAGAACTGGGC 2513
QY 2746 gaggctgtccagtggtgggattgcaacctggcgagcttaaccaaggcctggagaaggt 2805
DB 2514 GATGCTGTGCTCTTGGGGCTCCACCTGGTGGAGCTTACCAAGGGCTTGGACAGGCT 2573
QY 2806 ggaagagcctcagatggcatcagcaaaaagggaagcaccatgagcttggtgctctgggc 2865
DB 2574 GCAAAGCCTCGAGCGCGTGGCAAGGACGATCCACAGAGGSCCTCTGCTGCATGSGC 2633
QY 2866 ttccaggaggagccccctggggagagtggtgctgctgtccgggggacttoaagag 2925
DB 2634 TTCCAGGAGATCATGCTATCGGAAGAGCAGCAGCAGCTGCGGGGACTTCAAAACAG 2693
QY 2926 gaggagtggtgggtgaagagagagcagtggtggtgctgactgcccagagtgccc 2985
DB 2694 CAGGAGGACAGAGGGGGTGAAGAGAGGTGGTGGGTGCTGCAGTGGCCCGAGGTGGCC 2753
QY 2986 aaggctgaacctggtgggagcagcggcactgctgttccacggcagcttggggag 3045
DB 2754 AAGGCGCACCACTGGCTGGAGAGAGCGCCATGCTGTTCTTCCACTGACTTCGGGGAC 2813
QY 3046 ctccactgctgcacccacacagcaggaagagagacactggaagctgaggaggagactcc 3105
DB 2814 CTGCACTGTTGCCGCCCTTGGCAGGAAGAGAGACCTGGAAGCTGAAGAGAGTACTCT 2873
QY 3106 tcoctatgagctcctgggcagccacagcagagagcctggcagcagccgcgtgca 3165
DB 2874 TCTGTGTGTAACCTCTGGGAGCCCTGAGCAGAGGCCCCAGCCTCAGACCCCTGTGTC 2933
QY 3166 ccaagggccccactcatctcaccaagagagagtgagagtggtgctgagctccaagggc 3225
DB 2934 CCAAGGCCCCCTGATGTGCACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2993
QY 3226 ggtgggtgtcctgtgcaacctctcagggagtgccgttactgtgtgggcccactaagtg 3285
DB 2994 GGCTGGGTCTCTCCATGTCACCTCTCTGGGAGGCTGCTGTCTACTGGGGCCCTCTGTG 3053
QY 3286 ggcacagagtcacaggtccagagctggtttgagctcctctctctcacaacatgaagcaggt 3345
DB 3054 GGTGCCCAATCAAGAGGTTCAAGAGCTGGTTGAATCTCTCTGTGCACATGAAGCCAGGA 3113
QY 3346 gaagaggggctgagtgaggagcagctccagggatttccacacacctcggagcctctctg 3405
DB 3114 GAAGAGGGCCAGAGATGGAGAGGCTTCCAGGTAGTTCGCGCACCTCACAAGGCTCTCTG 3173
QY 3406 gccagaagcccaacagcctgctgtgccagggcgcccatcgcaaaagagagcctgtg 3465
DB 3174 GCCCGAAGCCTAACAAAGCCTGCTGTACCTGAGGGGCCCATGCTTAAGAAAGAGCCTGTG 3233
QY 3466 ccaaggggcaaaagccttacggagcctggtgacccgggggctgccagagccagggac 3525
DB 3234 CCACGGGGTAAGAGCTTAACGGAGCCGGGAGTACACCGGGGGGCTGCTGAGCGGGAAGAC 3293
QY 3526 tcccaatgagggcaccagtgctgcccacaaacacactctgtccctgaatcctgcacaggg 3585
DB 3294 TCTCCGTGACGGGTGCCAGCACTTCCAAAGACCTCTGTCTCCAGAGCTTGCACAGGG 3353
QY 3586 cccccccagggacagatggaagggctggagccccagggccgggggctgtggaagggctc 3645
DB 3354 CCCCCACAGGACAGCCGAGGGGGCTGGAGCCCAAGCCGCGGGGCTGTCTCAGAAAGGGCTC 3413
QY 3646 ccaagatgtactcgttctctcagggcctgagcccgactgagcccgacccgggacccca 3705
DB 3414 CCCAGATGTGTACCCGCTCGCTTACAGCTGTGAGTGAGCCCCAGACTCTCTGGACCTCCG 3473
QY 3706 ggcctgacacacacccctgacccccagacaaactgggggcaagcagcagcgagcgccctc 3765
DB 3474 GGCCTGACTTACCACCCCCACGCTCCCGACAACTGGGAGGTAAAGCAGGAGCTGCCCTC 3533
QY 3766 aagtcgggcaagcgggtggggaagcctcaccacagcctcctccagccccagcaacccg 3825
DB 3534 AAGTCTGGCAAGCGGTGCGAAACCATCACCAAGGCTGCATCCAGCCCCAGCAACCT 3593
QY 3826 gccgcccctgctgtgctccagcagcagcccgatggcctcccaagaccacgaagagacagac 3885
DB 3594 GCTGCCCTGCTGTGCTCTCAGACAGCAGCCCCATGGCTCCAAAGACCAAGAGGCCAGAC 3653
QY 3886 taccacagcgcctggcaagaccagcgtccatgatatccttgggtccacgacacaaacc 3945
DB 3654 TCTCCAGCATGCTGCGAAGACCAAGCCTCATGCTCTCGGCTTCAGGTTCACCAAAACC 3713
QY 3946 caggagatcttcaactcaagcggcgagccctcagggcgctcccaactgcccgt 4005
DB 3714 CAGCAGGTCTTCCATGTCAAAGCGCGGCGGCTCGGAGAGCGGATCCAGACTGCCGT 3773
QY 4006 gccaccaagagctcctcgacacagccactggccgcacattcaaggtctccagcagc 4065
DB 3774 GCCACCAAGAGCTCCCTGCTAACCACTTACCCTGCTGAGGTTCAGGTTCACAGTGGG 3833
QY 4065 ccccaagggggaggggtgagccagcgggcaaggttccccaaactggtgcagggcagc 4125
DB 3834 CCCCAGAGGAAGGTGCGATGAGCCAGCGGTCAAAGTACCCAGCTGCTGTACAGGAAC 3893
QY 4126 aagctctcagccgcccctccatgcctcaagaaagagcgccttcatgcccgcgtc 4185
DB 3894 AAGCTTTCAGATCGGCTCTCCACGCTCAAGAGAGTGGCTTTCATGGCAGCTGTC 3953
QY 4186 ccccaaggaagcggaacctgcttggcgacgagcagcagcagcagcagcagcagcagc 4245
DB 3954 CCGCCCAAAACAGGAGCTCATCTTGAGAAACAACA-----ATG 3993
QY 4246 geatgggggagatgggaagagagagagcctgaggttccccccactctctcaagagga 4305
DB 3994 GAAGTGGGGAGAGCGGAGGAGGAGGAGGCTGAGAGTTCCCTTGGGCTTACAGGAGA 4053
QY 4306 tgtcttcccaagaaagcccaagcccaagggcaatggcgagcctgccaacaaagctcc 4365
DB 4054 TGGCCTTACCCCAAGAGGGCCAGACCCAG---GGGTAGTGGGAGACCACCCCA---CCCC 4107
QY 4366 caccocccgagacccccattctcctcaagctcgccctc---gggagccttccaggggg 4423
DB 4108 CACCTTTGGAGCCACCTGCTGATGATGCTGAGGTGCTTACACAGTATCCTGCCCCAGTG 4167
QY 4424 ceatgaagacaaaggttctgcaccccccgaagggcgccgggctgaagctggaagcctc 4483
DB 4168 CTGTGAGGACCAAGGTGCTGCCCCGGAAGGGCGGCTCAAGCTGAGGAGCCATAG 4227
QY 4484 tgcagagatcacctcgcccagcctcagaggttgagctgtaagggcgccagggggcctc 4543


```
/gene="DKFZp434l139"
/Note="strong similarity to mouse MM1.1 product; The
frame shift was determined manually"
/codon_start=2
/product="hypothetical protein"
/protein_id="CA863768.1"
/db_xref="GI:6599192"

[translation="PCOPTRAOKPOGHTNYSYRKRKLTRAKNTSSPCKGRK
RRQQVLPDPAEPRIKLYTSSCRFSDSPFVREKRDATFTTCTVYNS
PGDAKPKRKSASSSSSSSFLDAAGASLAPGPSILOPRLPSLSTMLGP
VYSKALSTCLVCLCLONPANFKDLGDCGYPPEHCLPKKPKLKEKVRPEGTCEA
SLPLRTLLKPGCAAAATAGKPRDPADPAKQPLRTSARGLSRLOSCVCCDGR
DGEAAAPADKAGKHCSEKAPAEFGAEHCAVWTGTYLVLAGKFLGLEA
MKVAVMMSCSQAGATIGCCHKULHYTHPCASDAGCIFIEENFLSKCPKRRLP
"

polya_site 2601
BASE COUNT 547 a 843 c 780 g 490 t
ORIGIN

Query Match 16.1%; Score 970.4; DB 9; Length 2660;
Best Local Similarity 99.6%; Pred. No. 2.3e-161;
Matches 994; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 5028 gccctggccagccagagggccagagaaacagagccagccacacacacacagcagcta 5087
DB 1 GCCCTGCCAGCCCGACAGAGGGCCACAGAACAGCCAGGCGCACACCACTACAGCAGCTA 60
QY 5088 ttccagcggagcgctcaoctcgggcgggcgccgagaaacacacacacacacacacacac 5147
DB 61 TTCCAGCGGAGCGCTCCTCCTCGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 5148 gggcggtccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5207
DB 121 GGGCGGTGCCAAGCGAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
QY 5208 aatcgctcaagtaacattctcttgcagcggcgtgaggtcagcagcagcagcagcagcagc 5267
DB 181 AATCGCGCTCAAGTACATTCCTCTGTCAGCGGCGCTGAGGTGAGAGCGGCGGCGGCGG 240
QY 5268 ctctcaccctctgctgggtgagagcagcagcagcagcagcagcagcagcagcagcagcagc 5327
DB 241 CTCTCACCCTCTGCTGGGTGGAGAGCGGAGCGGCTTCCACCACATATGCACTGTGT 300
QY 5328 caactccctggatgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5387
DB 301 CAACTCCCTGGAGATGCGCCCAAGCCACAGCAAGAGCCTTCTCTGCTGCTGCTCTTC 360
QY 5388 ctcaactcgtctctctctctctctctctctctctctctctctctctctctctctctctct 5447
DB 361 CTCACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
QY 5448 aggtccatctctgagcggcgccctctctctctctctctctctctctctctctctctctct 5507
DB 421 AGCTCCATCTCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 5508 tbtgtttccagggccttgatgtaactcttctgttctgttctgttctgttctgttctgtt 5567
DB 481 TGTGTTTCCAGGCGCTGAGTACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
QY 5568 caacttcagagccttggagcactctgtggcctactactactactactactactactactact 5627
DB 541 CAACTTCAAGGACCTTGGGAGCCTCTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
QY 5628 aaagaagcc-aaactcaagagagaggtgcggcagagagcagcagcagcagcagcagcagc 5686
DB 601 AARGAGCCAAAACACTCAAGAGAGAGGTGCGGCCAGAGAGGACCTGTGAGAGGCGCTCG 660
QY 5687 tgcgcttgagagaactcaagtgctccagtgctgcagctgcgcgcgcgcgcgcgcgcgcgcgc 5746
DB 661 TGCGGCTTGAGAGAACACTCAAGAGGTCCCGAGTGTGAGGTGCGGCGGCGGCGGCGGAG 720
QY 5747 ccccccagg--tgagggccagctgaccggcgaagagagggccacactgctgagcagctgccc 5804
```

6.5%; Score 391.8; DB 11; Length 495;

Best Local Similarity -94.2%; Pred. No. 3.9e-59;
Matches 469; Conservative 1; Mismatches 21; Indels 7; Gaps 6;

```
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
```



```

match=match; cunas: Em:ABU006630 Em:U2U282 Em:U19345
/notice: Em:U2U282 Em:U2U282 Em:U2U282 Em:U2U282
Em:C75439 Em:C75439 Em:C75439 Em:C75439 Em:C75439 Em:C75439
Em:AA839615 Em:AA839615 Em:AA839615 Em:AA839615 Em:AA839615
Em:R08268 Em:R08268 Em:R08268 Em:R08268 Em:R08268 Em:R08268
Em:W17033 Em:W17033 Em:W17033 Em:W17033 Em:W17033 Em:W17033
Em:T04931 Em:T04931 Em:T04931 Em:T04931 Em:T04931 Em:T04931
/codon_start=-2
/evolution-not_experimental

```

[illegible]

Dg	5884	ATGCAGCATCGGGGAGCATATCCATGTGGGCCAAGCAGTGTGCGATTGGTTCACGAAGC	5222
Qy	1839	gaagagtaactcagacaagaagcaacttaatgccctgctcaattccctcccaagatgtcc	1698
Dd	5224	AAATATTTA-----CAATTATGCTCTACCTGATCCAAACCCCTCAATGATG	5177
Qy	1699	accagatctctgtgacaccacaggtctgcaactcgaagccctctgaagagaacagctccct	1758
Dd	5176	CCCATCTCT---AATTCATGATGTCAGAGCTTCAAAGGGTTGAGACTGAAAGGGATNCA	5120
Qy	1753	gggaacctgctctgcgactactagctctgcagagccctcgaagcgctgaaccttcaggtggag	1818
Dd	5119	GAAGAAGCAGCTGACAAATCTCGGTTGAGTGAATGATGCTGTGATCTCAAGTGAGCC	5060
Qy	1819	aacatctccacacccgctccagcagctctgtctctcccaagagctctgttcgcgcaagaagaa	1878
Dd	5059	AATCTTCCTAACTGTCACAGACATGTTACTTTCTATGCCCTGACTCCTCTCAGAGAG	5000
Qy	1879	ggtgcagaagaactctgtgc	1898
Dd	4999	ACCTCGAAGAGGCCCTCATC	4980

Query Match 1.8%; Score 110; DB 9; Length 43761;
Best Local Similarity 52.7%; Pred. No. 4.2e-10;
Matches 295; Conservative 0; Mismatches 250; Indels 15; Gaps 2;

[illegible]